Mon. Not. R. Astron. Soc. **394**, 1857–1874 (2009) Printed 15 July 2009 (MN LaTeX style file v2.2)

Understanding the H_2/HI Ratio in Galaxies

D. Obreschkow and S. Rawlings
Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, UK

Accepted 2009 January 12

ABSTRACT
We revisit the mass ratio $R_{\rm mol}^{\rm galaxy}$ between molecular hydrogen (H_2) and atomic hydrogen (HI) in different galaxies from a phenomenological and theoretical viewpoint. First, the local H_2-mass function (MF) is estimated from the local CO-luminosity function (LF) of the FCRAO Extragalactic CO-Survey, adopting a variable CO-to-H_2 conversion fitted to nearby observations. This implies an average H_2-density $\Omega_{\rm H_2} = (6.9\pm2.7)\cdot10^{-5}h^{-1}$ and $\Omega_{\rm H_2}/\Omega_{\rm HI} = 0.26\pm0.11$ in the local Universe. Second, we investigate the correlations between $R_{\rm mol}^{\rm galaxy}$ and global galaxy properties in a sample of 245 local galaxies. Based on these correlations we introduce four phenomenological models for $R_{\rm mol}^{\rm galaxy}$, which we apply to estimate H_2-masses for each HI-galaxy in the HIPASS catalog. The resulting H_2-MFs (one for each model for $R_{\rm mol}^{\rm galaxy}$) are compared to the reference H_2-MF derived from the CO-LF, thus allowing us to determine the Bayesian evidence of each model and to identify a clear best model, in which, for spiral galaxies, $R_{\rm mol}^{\rm galaxy}$ negatively correlates with both galaxy Hubble type and total gas mass. Third, we derive a theoretical model for $R_{\rm mol}^{\rm galaxy}$ for regular galaxies based on an expression for their axially symmetric pressure profile dictating the degree of molecularization. This model is quantitatively similar to the best phenomenological one at redshift $z = 0$, and hence represents a consistent generalization while providing a physical explanation for the dependence of $R_{\rm mol}^{\rm galaxy}$ on global galaxy properties. Applying the best phenomenological model for $R_{\rm mol}^{\rm galaxy}$ to the HIPASS sample, we derive the first integral cold gas-MF (HI+H_2+helium) of the local Universe.

Key words: ISM: atoms – ISM: molecules – ISM: clouds – radio lines: galaxies.

1 INTRODUCTION

The Interstellar Medium (ISM) plays a vital role in galaxies as their primordial baryonic component and as fuel or exhaust of stars. Hydrogen constitutes 74 per cent of the mass of the ISM. When it is cold and neutral it coexists in the atomic phase (HI) and molecular phase (H_2). While the former follows a smooth distribution across large galactic substructures, the latter is found in dense molecular clouds (Drapatz & Zinnecker 1984) acting as the sole crèches of newborn stars. The dissimilar but interlinked roles of HI and H_2 in substructure growth and star formation have caused a growing interest in simultaneous observations of both phases and cosmological simulations that distinguish between HI and H_2.

Extragalactic observations of HI often use its prominent 21-cm emission line, and currently comprise several thousand galaxies (HI Parkes All Sky Survey HIPASS, Barnes et al. 2001), and a maximum redshift of $z = 0.2$ (Verheijen et al. 2007). By contrast, most H_2-estimates must rely on indirect tracers, such as CO-lines, with uncertain conversion factors. Consequently, the phase ratio of neutral hydrogen $R_{\rm mol} \equiv {\rm d}M_{\rm H_2}/{\rm d}M_{\rm HI}$ and its value for entire galaxies $R_{\rm mol}^{\rm galaxy} \equiv M_{\rm H_2}/M_{\rm HI}$ remain debated, and estimates of the universal density ratio $R_{\rm mol}^{\rm universe} \equiv \Omega_{\rm H_2}/\Omega_{\rm HI}$ vary by an order of magnitude in the local Universe (e.g. 0.14, 0.42, 1.1 stated respectively by Boselli et al. 2002, Keres et al. 2003, Fukugita et al. 1998).

Ultimately, the uncertainties of H_2-measurements hinder the reconstruction of cold gas masses $M_{\rm gas} = (M_{\rm HI} + M_{\rm H_2})/\beta$, where $\beta \approx 0.74$ is the standard fraction of hydrogen in neutral gas with the rest consisting of helium (He) and a minor fraction of heavier elements. The limitations of comparing $M_{\rm HI}$ to $M_{\rm gas}$ caused by the measurement uncertainties of $M_{\rm H_2}$ culminate in severe difficulties to compare statistically tight cold gas-mass functions (MFs) of modern cosmological simulations with precise HI-MFs extracted from HI-surveys, such as HIPASS. Both simulations and surveys have reached statistical accuracies far better than any current model for $R_{\rm mol}^{\rm galaxy}$, and hence the comparison of observations with simulations is mainly limited by the uncertainty of $R_{\rm mol}^{\rm galaxy}$.

As an illustration, Fig. 1 displays the observed HI-MF from the HIPASS sample (Zwaan et al. 2005) together with several simulated HI-MFs. The latter are based on the cold gas masses of the simulated galaxies produced by two different galaxy formation models applied to the Millennium Simulation (Bower et al. 2006; De Lucia & Blaizot 2007). We have converted these cold gas masses into HI-masses using four models for $R_{\rm mol}^{\rm galaxy}$ from the literature (Young & Knezek 1989; Keres et al. 2003; Boselli et al.

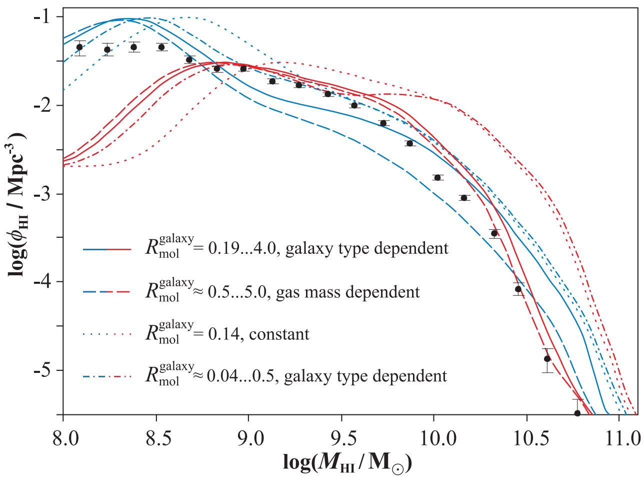

Figure 1. Dots represent the observed HI-MF by Zwaan et al. (2005); lines represent simulated HI-MFs derived from the semi-analytic models by Bower et al. (2006, red lines) and De Lucia & Blaizot (2007, blue lines). The four models of $R_{\rm mol}^{\rm galaxy}$ were adopted or derived from Young & Knezek (1989, solid lines), Keres et al. (2003, dashed lines), Boselli et al. (2002, dotted lines), and Sauty et al. (2003, dash-dotted lines).

2002; Sauty et al. 2003). The figure adopts the Hubble constant of the Millennium Simulation, i.e. $h = 0.73$, where h is defined by $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$ with H_0 being the present-day Hubble constant. The differential gas density $\phi_{\rm HI}$ is defined as $\phi_{\rm HI} \equiv {\rm d}\rho_{\rm HI}/{\rm d}\log M_{\rm HI}$, where $\rho_{\rm HI}(M_{\rm HI})$ is the space density (i.e. number per volume) of HI-sources of mass $M_{\rm HI}$. In Fig. 1 different models for galaxy formation are distinguished by colour, while the models of $R_{\rm mol}^{\rm galaxy}$ are distinguished by line type. Clearly, any conclusion regarding the two galaxy formation models based on their HI-MFs is affected by the choice of the model for $R_{\rm mol}^{\rm galaxy}$.

This paper presents a state-of-the-art analysis of the galaxy-dependent phase ratio $R_{\rm mol}^{\rm galaxy}$, the H_2-MF and the integral cold gas-MF (HI+H_2+He), utilizing various observational constraints. In Section 2, the determination of H_2-masses via CO-lines is revisited and an empirical, galaxy-dependent model for the CO-to-H_2 conversion factor (X-factor) is derived from direct measurements of a few nearby galaxies (Boselli et al. 2002 and references therein). In Section 3, this model is applied to recover an *H_2-MF from the CO-luminosity function (LF)* by Keres et al. (2003). The resulting H_2-MF significantly differs from the one obtained by Keres et al. (2003) using a constant X-factor. Section 4 presents an independent derivation of the *H_2-MF from a HI-sample* with well characterized sample completeness (HIPASS, Barnes et al. 2001). This approach is less prone to completeness errors, but it premises an estimate of the H_2/HI-mass ratio $R_{\rm mol}^{\rm galaxy}$. Therefore, we propose four phenomenological models of $R_{\rm mol}^{\rm galaxy}$ (as functions of other galaxy properties) and compute their Bayesian evidence by comparing the resulting H_2-MFs to the reference H_2-MF derived from the CO-LF. This empirical method is supported by Section 5, where we analytically derive a galaxy-dependent model for $R_{\rm mol}^{\rm galaxy}$ on the basis of the relation between $R_{\rm mol}$ and the pressure of the ISM (Leroy et al. 2008). A brief discussion and a derivation of an integral cold gas-MF (HI+H_2+He) are presented in Section 6. Section 7 concludes the paper with a summary and outlook.

2 THE VARIABLE CO-TO-H_2 CONVERSION

2.1 Background: basic mass measurement of HI and H_2

HI emits rest-frame 1.42 GHz radiation ($\lambda = 0.21$ m) originating from the hyperfine spin-spin relaxation. Especially cold HI ($T \sim 50-100$ K, see Ferrière 2001) also appears in absorption against background continuum sources or other HI-regions, but makes up a negligible fraction in most galaxies. Within this assumption, HI can be considered as optically thin on galactic scales, and hence the HI-line intensity is a proportional mass tracer,

$$\frac{M_{\rm HI}}{\rm M_\odot} = 2.36\cdot 10^5\cdot\frac{S_{\rm HI}}{\rm Jy\,km\,s^{-1}}\cdot\left(\frac{D_{\rm l}}{\rm Mpc}\right)^2, \qquad (1)$$

where $S_{\rm HI}$ is the integrated HI-line flux density and $D_{\rm l}$ is the luminosity distance to the source.

Unlike HI-detections, direct detections of H_2 in emission rely on weak lines in the infrared and ultraviolet bands (Dalgarno 2000) and have so far been limited to the Milky Way and a few nearby galaxies (e.g. Valentijn & van der Werf 1999). Occasionally, H_2 has also been detected at high redshift ($z\approx$2–4) through absorptions lines associated with damped Lyman α systems (Ledoux et al. 2003; Noterdaeme et al. 2008). All other H_2-mass estimates use indirect tracers, mostly rotational emission lines of carbon monoxide (CO) – the second most abundant molecule in the Universe. The most frequently used CO-emission line stems from the relaxation of the $J = 1$ rotational state of the predominant isotopomer $^{12}C^{16}O$. Radiation from this transition is referred to as CO(1–0)-radiation and has a rest-frame frequency of 115 GHz ($\lambda = 2.6\cdot 10^{-3}$ m), detectable with millimeter telescopes. The conversion between CO(1–0)-radiation and H_2-masses is very subtle and generally expressed by the X-factor,

$$X \equiv \frac{N_{\rm H_2}/{\rm cm^{-2}}}{I_{\rm CO}/({\rm K\,km\,s^{-1}})}\cdot 10^{-20}, \qquad (2)$$

where $N_{\rm H_2}$ is the column density of molecules and $I_{\rm CO}$ is the integrated CO(1–0)-line intensity per unit surface area defined via the surface brightness temperature T_ν in the Rayleigh-Jeans approximation. Explicitly, $I_{\rm CO} \equiv \int T_\nu {\rm d}V = \lambda\int T_\nu{\rm d}\nu$, where V is the radial velocity, ν is the frequency, and $\lambda = |{\rm d}V/{\rm d}\nu|$ is the wavelength. This definition of the X-factor implies a mass-luminosity relation analogous to Eq. (1) (see review by Young & Scoville 1991),

$$\frac{M_{\rm H_2}}{\rm M_\odot} = 580\cdot X\cdot\left(\frac{\lambda}{\rm mm}\right)^2\cdot\frac{S_{\rm CO}}{\rm Jy\,km\,s^{-1}}\cdot\left(\frac{D_{\rm l}}{\rm Mpc}\right)^2, \qquad (3)$$

where $S_{\rm CO} \equiv \int S_{{\rm CO},\nu}{\rm d}V$ denotes the integrated CO(1–0)-line flux and $D_{\rm l}$ the luminosity distance. $S_{{\rm CO},\nu}$ is the flux density per unit frequency, for example expressed in Jy, and thus $S_{\rm CO}$ has units like ${\rm Jy\,km\,s^{-1}}$. Note that $S_{\rm CO}$ relates to the physical flux F, defined as power per unit surface, via a factor λ, i.e. $F \equiv \int S_{{\rm CO},\nu}{\rm d}\nu = \lambda^{-1}S_{\rm CO}$. CO-luminosities are often defined as $L_{\rm CO} \equiv 4\pi D_{\rm l}^2 S_{\rm CO}$ (giving units like ${\rm Jy\,km\,s^{-1}}\,(h^{-1}\,{\rm Mpc})^2$), thus relating to actual radiative power $P_{\rm CO}$ via $P_{\rm CO} = \lambda^{-1}L_{\rm CO}$. In the λ-dependent notation above, Eq. (3) remains valid for other molecular emission lines, as long as the X-factor is redefined with the respective intensities in the denominator of Eq. (2).

2.2 Variation of the X-factor among galaxies

The theoretical and observational determination of the X-factor is a highly intricate task with a long history, and it is perhaps one of the biggest challenges for future CO-surveys.

Theoretically, the difficulty to estimate X arises from the indirect mechanism of CO-emission and from the optical thickness of CO(1–0)-radiation. CO resides inside molecular clouds along with H_2 and acquires rotational excitations from H_2-CO collisions, which can subsequently decay via photon-emission. This mechanism implies that the CO(1–0)-luminosity per unit molecular mass a priori depends on three aspects: (i) the amount of CO per unit H_2, i.e. the CO/H_2-mass ratio; (ii) the thermodynamic state variables dictating the level populations of CO; (iii) the geometry of the molecular region influencing the degree of self-absorption.

The reason why the CO-luminosity can be used at all as a H_2-mass tracer is a statistical one. In fact, CO-luminosities are normally integrated over kiloparsec or larger scales, such as is inevitable given the spatial resolution of most extragalactic CO-surveys. Therefore, hundreds or thousands of molecular clouds are combined into one measurement, and cloud properties, such as geometries and thermodynamic state variables, probably tend towards a constant average, as long as most lines-of-sight to individual clouds do not pass through other clouds, where they would be affected by self-absorption. The latter assumption seems correct for all but nearly edge-on spiral galaxies (Ferrière 2001; Wall 2006). It is hence likely that the different geometries and thermodynamic variables of molecular clouds can be neglected in the variations of X and we expect X to depend most significantly on the average CO/H_2-mass ratio of the considered galaxy or galaxy part. However, the determination of the CO/H_2-ratio is itself difficult and its relation to the overall metallicity of the galaxy is uncertain.

Observational estimations of X require CO-independent H_2-mass measurements, which are limited to the Milky Way and a few nearby galaxies. Typical methods use the virial mass of giant molecular clouds assumed to be completely molecularized (Young & Scoville 1991), the line ratios of different CO-isotopomers (Wild et al. 1992), mm-radiation from cold dust associated with molecular clouds (Guelin et al. 1993), and diffuse high energy γ-radiation caused by interactions of cosmic-rays with the ISM (Bertsch et al. 1993; Hunter et al. 1997).

Early measurements suggested a fairly constant X in the inner $2-10$ kpc of the Galaxy, leading several authors to the conclusion that X does not significantly depend on cloud properties and metallicity (e.g. Young & Scoville 1991). This finding has recently been supported by Blitz et al. (2007), who analyzed five galaxies in the local group and found no clear trend between metallicity and X. The results of Young & Scoville (1991) and Blitz et al. (2007) rely on the assumption that molecular clouds are virialized. Using the same method Arimoto et al. (1996) detected strong variations of X amongst galaxies and galactic substructures, and they found the empirical power-law relation $X \propto (\mathrm{O/H})^{-1}$. Israel (2000) pointed out that molecular clouds cannot be considered as virialized structures, and using far-infrared measurements rather than the virial theorem, Israel (1997) found an even tighter and steeper relation in a sample of 14 nearby galaxies, $X \propto (\mathrm{O/H})^{-2.7}$.

In summary, despite rigorous efforts to measure X and its relation to metallicity, the empirical findings remain uncertain and depend on the method used to measure X. Since we cannot overcome this issue, we shall use a model for X that relies on different methods to measure X, such as presented by Boselli et al. (2002). Their sample consists of 14 nearby galaxies covering an order of magnitude in O/H-metalicity. This sample includes early- and late-type spiral galaxies, as well as irregular objects and starbursts. For these galaxies X was determined from three different methods: the virial method, mm-data, and γ-ray data. Their data varies from $X = 0.88$ in the center of the face-on Sbc-spiral galaxy M 51 to $X \approx 60$ in NGC 55, a barred irregular galaxy seen edge-on. The high values ($X \gtrsim 10$) are often associated with dwarf galaxies and nearly edge-on spiral galaxies, thus consistent with the interpretation of increased CO(1–0) self-absorption in these objects. Typical values for non-edge-on galaxies lie around $X \approx 1-5$.

Model for log(X)	c_0	c_1	rms	$\ln B$
c_0	0.43 ± 0.15	-	0.45	0.0
$c_0 + c_1 \cdot \log(\mathrm{O/H})$	-2.90 ± 0.20	-1.02 ± 0.05	0.19	5.1
$c_0 + c_1 \cdot (M_\mathrm{B} - 5 \log h)$	3.67 ± 0.25	0.176 ± 0.006	0.29	3.3
$c_0 + c_1 \cdot \log(L_\mathrm{CO})$	1.85 ± 0.15	-0.288 ± 0.05	0.29	2.5

Table 1. Comparison of different models for the X-factor: c_0 and c_1 are the best parameters (Gaussian errors are coupled), rms is the rms-deviation of the data from the model, and B is the Bayes factor of each model with respect to the constant model (first row).

For the particular data set of Boselli et al. (2002), we shall check the validity of a constant-X model against variable models for X, by comparing their Bayesian evidence – a powerful tool for model selection (e.g. Sivia & Skilling 2006). The underlying idea is that the probability $p(M|d)$ of a model M given the data set d is proportional to the probability $p(d|M)$ of d given M, provided the compared models are a priori equally likely (Bayes theorem). The probability $p(d|M)$ is also called the *Bayesian evidence* and can be computed as,

$$p(d|M) = \int_\Omega p(d|\theta, M)\pi(\theta|M)\mathrm{d}\theta \qquad (4)$$

where θ denotes the vector of free parameters of model M and Ω the corresponding parameter space; $p(d|\theta, M)$ designates the probability of the data given a parameter choice θ and it typically includes measurement uncertainties of the data. The prior knowledge on the parameters is encoded in the probability density function $\pi(\theta|M)$, which satisfies the normalization condition $\int_\Omega \pi(\theta|M)\mathrm{d}\theta = 1$. Two competing models M_1 and M_2 are compared by their odds, commonly referred to as the *Bayes factor* $B \equiv p(d|M_1)/p(d|M_2)$. According to Jeffrey's scale (Jeffreys 1961) for the strength of evidence, $|\ln B| < 1$ is *inconclusive*, while $|\ln B| = 1$ reveals *positive evidence* in favour of model M_1 (probability=0.750), $|\ln B| = 2.5$ depicts *moderate evidence* (probability=0.923), and $|\ln B| = 5$ expresses *strong evidence* (probability=0.993).

We consider the four models listed in Table 1: a constant model, where $\theta = (c_0)$, and three linear models, where $\theta = (c_0, c_1)$. The data are a sample of 14 nearby galaxies, for which X was measured (Table 2); X-factors and O/H-metallicities are taken from Boselli et al. (2002) and references therein, while M_B-magnitudes were taken from the HyperLeda database (Paturel et al. 2003), and CO(1–0)-luminosities L_CO were derived from the references indicated in Table 2.

For practical purposes we limit the parameter space Ω to $c_0 \in [-10, 10]$ and $c_1 \in [-2, 2]$ and take the prior probabilities as homogeneous within Ω, i.e. $\pi(\theta|M) = 1/|\Omega|$. The probability $p(d|\theta, M)$ in Eq. (4) is calculated as the product,

$$p(d|\theta, M) = \prod_\mathrm{i} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{[\log(X_\mathrm{i}^\mathrm{data}) - \log(X_\mathrm{i}^\mathrm{model})]^2}{2\sigma^2}\right\} \qquad (5)$$

where i labels the different galaxies listed in Table 2 and σ denotes the measurement uncertainty of log(X). We set σ equal the average value $\sigma = 0.13$, for all 14 galaxies. (In fact adopting the specific σ-values listed in Table 2 leads to very similar results, but could be potentially dangerous as the small value $\sigma = 0.01$ of the Milky Way is likely underestimated.)

Object	log(O/H) [(a)]	$M_{\mathrm{B}}^{(b)}$ $-5\log h$	$\log(L_{\mathrm{CO}})$	log(X) [(a)]
SMC	-3.96	-16.82	-2.04 [(c)]	1.00
NGC1569	-3.81	-15.94	-1.60 [(d)]	1.18
M31	-2.99	-20.23	-1.40 [(e)]	0.38 ± 0.21
IC10	-3.69	-15.13	-1.09 [(f)]	0.82 ± 0.12
LMC	-3.63	-17.63	-0.68 [(g)]	0.90
M81	-3	-19.90	-0.07 [(h)]	-0.15
M33	-3.22	-18.61	0.20 [(i)]	0.70 ± 0.11
M82	-3	-17.30	0.67 [(d)]	0.00
NGC4565	-	-21.74	1.12 [(h)]	0.00
NGC6946	-2.94	-20.12	1.24 [(h)]	0.26
NGC891	-	-19.43	1.48 [(h)]	0.18
M51	-2.77	-19.74	1.80 [(h)]	-0.22
Milky Way	-3.1	-19.63	-	0.19 ± 0.01
NGC6822	-3.84	-16.07	-	0.82 ± 0.20

Table 2. Observational data used for the derivation of a variable X-factor (Section 2.2). L_{CO} is given in units of Jy km s^{-1} (h^{-1} Mpc)^2. (a) O/H-metallicities and X-factors from Boselli et al. (2002), (b) absolute, extinction-corrected B-Magnitudes from the HyperLeda database (Paturel et al. 2003), (c) Rubio et al. (1991), (d) Young et al. (1989), (e) Heyer et al. (2000), (f) Leroy et al. (2006), (g) Fukui et al. (1999), (h) Sage (1993), (i) Heyer et al. (2004).

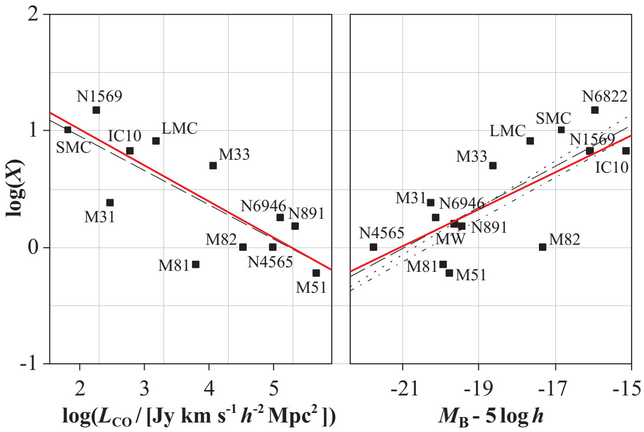

Figure 2. Points represent observed X-factors as a function of CO(1–0)-power L_{CO} and absolute blue magnitude M_{B} for 14 local galaxies. Red solid lines represent linear regressions respecting the mutual relation between L_{CO} and M_{B} given in Eq. (6); dashed lines represent independent linear regressions; the dotted line represents the linear fit found by Arimoto et al. (1996); and the dash-dotted line represents the linear fit found by Boselli et al. (2002).

The evidence integrals were solved numerically using a Monte Carlo sampling of the parameter space. The resulting Bayes factors (listed in Table 1) reveal moderate to strong Bayesian evidence for a variable X-factor given the X-factors presented by Boselli et al. (2002). Among the different variable models for log(X), the best one depends linearly on log(O/H) (highest Bayes factor), as expected from the natural dependence of the CO/H_2 ratio on the O/H ratio. However, log(X) is also well correlated with M_{B} and $\log(L_{\mathrm{CO}})$, and hereafter we will use those relations because of the widespread availability of M_{B} and L_{CO} data. In fact, a X-factor depending on L_{CO} simply translates to a non-linear conversion of CO-luminosities into H_2-masses. If the two linear regressions between log(X) and M_{B} and between log(X) and $\log(L_{\mathrm{CO}})$ were determined independently, they would imply a third linear relation between M_{B} and $\log(L_{\mathrm{CO}})$. The latter can, however, be determined more accurately from larger galaxy samples. The sample presented in Section 4.1 (245 galaxies) yields

$$\log(L_{\mathrm{CO}}) \approx -4.5 - 0.52\,(M_{\mathrm{B}} - 5\log h), \tag{6}$$

where L_{CO} is taken in units of Jy km s^{-1} (h^{-1} Mpc)^2. To get the best result, we imposed this relation, while simultaneously minimizing the square deviations of the two regressions between log(X) and respectively M_{B} and $\log(L_{\mathrm{CO}})$. In such a way we find

$$\log(X) = 1.97 - 0.308\,\log(L_{\mathrm{CO}}) \pm \sigma_{\mathrm{X}}, \tag{7}$$

$$\log(X) = 3.36 + 0.160\,(M_{\mathrm{B}} - 5\log h) \pm \sigma_{\mathrm{X}}. \tag{8}$$

These two relations are shown in Fig. 2 (red solid lines). For comparison, the independent regressions, obtained without imposing the relation given in Eq. (6), are plotted as dashed lines. These relations correspond to the parameters c_0 and c_1 given in Table 1. Other regressions found by Arimoto et al. (1996) and Boselli et al. (2002) are also displayed. Their approaches are similar, but Arimoto et al. (1996) used less galaxies (8 instead of 14). The 14 data points in Fig. 2 are scattered around the relations of Eqs. (7) and (8) with the same rms-deviation of 0.29 in log(X). Combined with the average measurement uncertainty of $\sigma = 0.13$, this gives an estimated true physical scatter in log(X) of $\sigma_{\mathrm{X}} = (0.29^2 - 0.13^2)^{1/2} = 0.26$.

The variable models of X given in Eqs. (7) and (8) will be applied in Sections 3 and 4. In order to account for the uncertainties of X highlighted in the beginning of this section, we shall also present the results for a constant X-factor with random scatter in Section 3.

3 DERIVING THE H_2-MF FROM THE CO-LF

Using the variable model for the X-factor of Eq. (7), we shall now recover the local H_2-mass function (H_2-MF) from the CO-LF presented by Keres et al. (2003). The latter is based on a far infrared-selected subsample of 200 galaxies from the FCRAO Extragalactic CO-Survey (Young et al. 1995), which successfully reproduced the 60μm-LF, thus limiting the errors caused by the incompleteness of the sample. Keres et al. (2003) themselves derived a H_2-MF using a constant model $X = 3$, which probably leads to an overestimation of the H_2-abundance, especially in the high mass end, where the X-factors tend to be lower according to the data shown in Section 2.2.

We applied Eq. (7) with scatter $\sigma_{\mathrm{X}} = 0.26$ to the individual data points of the CO-LF given by Keres et al. (2003). The resulting H_2-MF – hereafter the *reference H_2-MF* – is shown in Fig. 3 together with the *original H_2-MF* derived by Keres et al. (2003) using the constant factor $X = 3$ without scatter. To both functions we fitted a Schechter function (Schechter 1976) of the form

$$\phi_{\mathrm{H_2}} = \ln(10)\cdot\phi^*\cdot\left(\frac{M_{\mathrm{H_2}}}{M^*}\right)^{\alpha+1}\exp\left[-\left(\frac{M_{\mathrm{H_2}}}{M^*}\right)\right] \tag{9}$$

by minimizing the weighted square deviations of all but the highest H_2-mass bin. Keres et al. (2003) argue that this bin may contain a CO-luminous subpopulation of starburst galaxies, similarly to the situation in the far infrared continuum (Yun et al. 2001). In any case the last bin only marginally contributes to the universal H_2-density. The Schechter function parameters are given in Table 3, as well as the reduced χ^2 of the fits, total H_2-densities $\rho_{\mathrm{H_2}}$ and $\Omega_{\mathrm{H_2}} \equiv \rho_{\mathrm{H_2}}/\rho_{\mathrm{crit}}$, and the average molecular ratio $R_{\mathrm{mol}}^{\mathrm{universe}} \equiv \Omega_{\mathrm{H_2}}/\Omega_{\mathrm{HI}}$. Both $\rho_{\mathrm{H_2}}$ and $\Omega_{\mathrm{H_2}}$ were evaluated from the fitted Schechter function rather than the binned data, and $\Omega_{\mathrm{HI}} = (2.6 \pm 0.3)\,h^{-1}10^{-4}$ was adopted from the HIPASS analysis by Zwaan et al. (2005).

Our new reference H_2-MF is compressed in the mass-axis

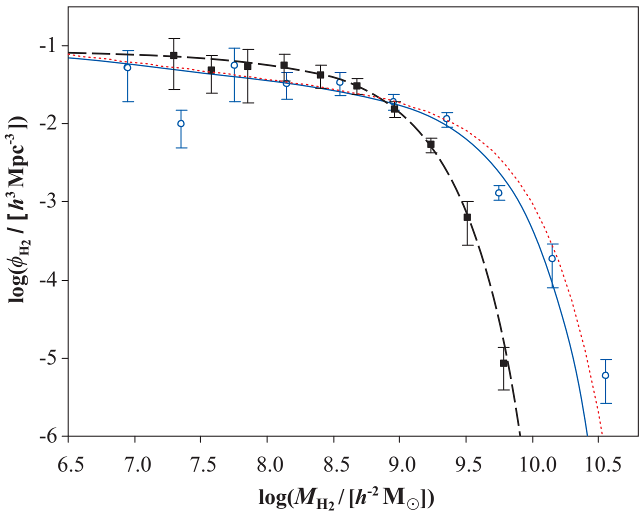

Figure 3. Filled squares represent our reference H_2-MF derived directly from the observed CO-LF (Keres et al. 2003) using the variable X-factor of Eq. (7) with scatter $\sigma_X = 0.26$. Open circles are the original H_2-MF obtained by Keres et al. (2003) using a constant factor $X = 3$ without scatter. The dashed and solid lines represent Schechter function fits to our reference H_2-MF and the original H_2-MF, while the dotted line represents the Schechter function corresponding to a constant X-factor $X = 3$ with scatter σ_X.

compared to the original one, and our estimate of ρ_{H_2} (Table 3) is 33 per cent smaller. The global H_2/HI-mass ratio drops to 0.26 ± 0.11, implying a total cold gas density of $\Omega_{gas} = (4.4 \pm 0.8) \cdot 10^{-4}\ h^{-1}$. The composition of cold gas becomes: 59 ± 6 per cent HI, 15 ± 6 per cent H_2, 26 per cent He and metals, where the uncertainties of HI and H_2 are anti-correlated.

It is interesting to observe the quality of the Schechter function fits: the fit to our reference H_2-MF is much better than the one to the original H_2-MF (Keres et al. 2003). Since the original MF is a simple shift of the CO-LF (constant X-factor), the Schechter function fit to our reference H_2-MF is also much better than the fit to the CO-LF. We could demonstrate that this difference is partially caused by the scatter $\sigma_X = 0.26$, applied to the variable X-factor when deriving the reference H_2-MF from the CO-LF. Scatter averages the densities in neighboring mass bins, hence smoothing the reference MF. Additionally, there is a fundamental reason for the rather poor Schechter function fit of the CO-LF: It is formally impossible to describe both the H_2-MF and the CO-LF with Schechter functions, if the two are interlinked via the linear transformation of Eq. (7). Yet, in analogy to the HI-MF (Zwaan et al. 2005), it is likely that the H_2-MF is well matched by a Schechter function, hence implying that the CO-LF deviates from a Schechter function.

We finally note, that the faint end of the reference H_2-MF is nearly flat (i.e. $\alpha = -1$), such that the total H_2-mass is dominated by masses close to the Schechter function break at $M^* \approx 10^9 M_\odot$. In particular, the faint end slope is flatter than for the HI-MF, where $\alpha = -1.37$ (Zwaan et al. 2005), but it should be emphasized that this does not imply that small cold gas masses have a lower molecular fraction. In fact, the contrary is suggested by the observations shown in the Section 4.

For completeness, we re-derived the H_2-MF from the CO-LF using a constant X-factor $X = 3$ (like Keres et al. 2003) with the same Gaussian scatter $\sigma_X = 0.26$ as used for our variable model of X. The best Schechter fit for the resulting H_2-MF is also displayed in Fig. 3. The difference between this H_2-MF and the original H_2-MF by Keres et al. (2003) demonstrates that the scatter of X stretches the high mass end towards higher masses.

	reference H_2-MF (variable X)	original H_2-MF (constant X)
M^*	$7.5 \cdot 10^8\ h^{-2}\ M_\odot$	$2.81 \cdot 10^9\ h^{-2}\ M_\odot$
α	-1.07	-1.18
ϕ^*	$0.0243\ h^3\ \mathrm{Mpc}^{-3}$	$0.0089\ h^3\ \mathrm{Mpc}^{-3}$
Red. χ^2	0.05	2.55
ρ_{H_2}	$(1.9 \pm 0.7) \cdot 10^7\ h\ M_\odot\ \mathrm{Mpc}^{-3}$	$(2.8 \pm 1.1) \cdot 10^7\ h\ M_\odot\ \mathrm{Mpc}^{-3}$
Ω_{H_2}	$(0.69 \pm 0.27) \cdot 10^{-4}\ h^{-1}$	$(1.02 \pm 0.39) \cdot 10^{-4}\ h^{-1}$
$R_{mol}^{universe}$	0.26 ± 0.11	0.39 ± 0.16

Table 3. Schechter function parameters, reduced χ^2, and universal mass densities as obtained by integrating the Schechter functions. $R_{mol}^{universe} \equiv \Omega_{H_2}/\Omega_{HI}$ is the global H_2/HI-mass ratio of the local Universe. The very small reduced χ^2 of our reference H_2-MF arises from a spurious smoothing introduced by the scatter σ_X.

4 PHENOMENOLOGICAL MODELS FOR THE H_2/HI-MASS RATIO

In this section, we shall introduce four *phenomenological* models for the H_2/HI-mass ratio R_{mol}^{galaxy} of individual galaxies. Each model will be used to recover a H_2-MF from the HIPASS HI-catalog (Barnes et al. 2001), thus demonstrating an alternative way to determine the H_2-MF to the CO-based approach. Comparing the H_2-MFs of this section with the reference H_2-MF derived from the CO-LF (Section 3) will allow us to determine the statistical evidence of the models for R_{mol}^{galaxy}.

4.1 Observed sample

The sample of galaxies used in this section is presented in Appendix A and consists of 245 distinct objects with simultaneous measurements of integrated HI-line fluxes and CO(1–0)-line fluxes. The latter were drawn from 9 catalogs in the literature, and, where not given explicitly, recomputed from indicated H_2-masses by factoring out the different X-factors used by the authors. HI line fluxes were taken from HIPASS via the optical cross-match catalog HOPCAT (Doyle et al. 2005). Both line fluxes were homogenized using h-dependent units, where they depend on the Hubble parameter h. Additional galaxy properties were adopted from the homogenous reference database “HyperLeda” (Paturel et al. 2003). These properties include numerical Hubble types T, extinction corrected blue magnitudes M_B, and comoving distances D_l corrected for Virgo infall. In the few cases, where these properties were unavailable in the reference catalog, they were copied from the original reference for CO-fluxes. For each galaxy we calculated HI- and H_2-masses using respectively Eqs. (1) and (3). The variable X-factors were determined from the blue magnitudes according to Eq. (8). We chose to compute X from M_B rather than from L_{CO}, because of the smaller measurement uncertainties of the M_B data. Finally, total cold gas masses $M_{gas} = (M_{HI} + M_{H_2})/\beta$ and mass ratios $R_{mol}^{galaxy} = M_{H_2}/M_{HI}$ were calculated for each object. While the masses depend on the distances and hence on the Hubble parameter h, the mass ratios $R_{mol}^{galaxy} = M_{H_2}/M_{HI}$ are independent of h.

This sample covers a wide range of galaxy Hubble types, masses, and environments, and has 49 per cent overlap with the subsample of the FCRAO Extragalactic CO-Survey used for the derivation of the reference H_2-MF in Section 3. We deliberately

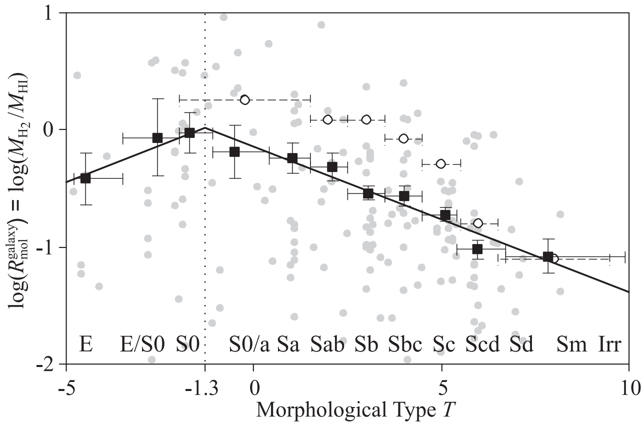

Figure 4. H_2/HI-mass ratio versus numerical Hubble type T. Grey dots represent the empirical data obtained by applying the variable X-factor of Eq. (8) with scatter to the CO-measurements. Black points represent the binned data; vertical bars represent statistical uncertainties obtained via bootstrapping, i.e. they depict a 1-σ confidence interval of the bin average obtained by examining 10^4 random half-subsets of the full data; horizontal bars represent the bin intervals. The solid line represents model 1 fitted to the data points. Open circles and dashed bars denote the binned data of the original paper by Young & Knezek (1989).

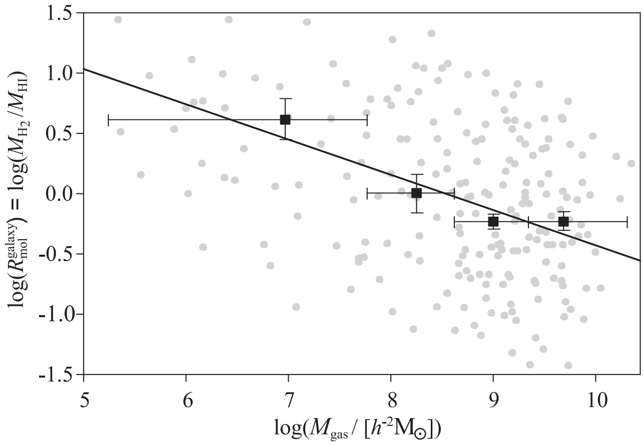

Figure 5. H_2/HI-mass ratio versus total cold gas mass $M_{\rm gas} \equiv (M_{\rm HI} + M_{\rm H_2})/\beta$. Grey dots represent the empirical data obtained by applying the variable X-factor of Eq. (8) with scatter to the CO-measurements. Black points represent the binned data; vertical bars represent the 1-σ confidence intervals; horizontal bars represent the bin intervals. The solid line represents model 2 fitted to the data points.

limited the sample overlap to 50 per cent in order to control possible sample biases.

We emphasize that this sample exhibits unknown completeness properties, which a priori presents a problem for any empirical model for $R_{\rm mol}^{\rm galaxy}$. However, as long as a proposed model is formally complete in the sense that it embodies the essential correlations with a set of free parameters, these parameters can be determined accurately even with an incomplete set of data points. The difficulty in the present case is that no reliable complete model for the molecular fraction $R_{\rm mol}^{\rm galaxy}$ has yet been established. We shall bypass this issue by proposing several models for $R_{\rm mol}^{\rm galaxy}$ that will be verified with hindsight (Section 4.2). Additional verification will become possible in Section 5, where we shall derive a physical model for $R_{\rm mol}^{\rm galaxy}$.

4.2 Phenomenological models for $R_{\rm mol}^{\rm galaxy}$

The galaxy sample of Section 4.1 reveals moderate correlations between $R_{\rm mol}^{\rm galaxy}$ and respectively T, $M_{\rm gas}$ and $M_{\rm B}$. These correlations motivate the models proposed below. Other correlations were looked at, such as a correlation between $R_{\rm mol}^{\rm galaxy}$ and environment, which may be suspected from stripping mechanisms acting differently on HI and H_2. However no conclusive trends could be identified given the observational scatter of $R_{\rm mol}^{\rm galaxy}$. All our models are first presented with free parameters, which are fitted to the data at the end of this section.

Model 0 ($R_{\rm mol,0}^{\rm galaxy}$) assumes a constant H_2/HI-ratio $R_{\rm mol}^{\rm galaxy}$, such as is often used in the literature,

$$\log(R_{\rm mol,0}^{\rm galaxy}) = q_0 + \sigma_{\rm phy,0}, \tag{10}$$

where q_0 is a constant and $\sigma_{\rm phy,0}$ denotes an estimate of the physical scatter of perfectly measured data relative to the model.

Model 1 is galaxy-type dependent, as suggested by earlier studies revealing a trend for $R_{\rm mol}^{\rm galaxy}$ to increase from late-type spiral galaxies to early-type ones (e.g. Young & Knezek 1989; Sauty et al. 2003). The type dependence of our sample is displayed in Fig. 4. The binned data clearly show a monotonic increase of the molecular fraction by roughly an order of magnitude when passing from late-type spiral galaxies (Scd–Sd) to early-type spiral and lenticular galaxies (S0–S0/a). The unbinned data illustrate the importance of parameterizing the physical scatter. The Hubble type dependence can be widely explained by the effect of the bulge component on the disc size, as detailed in Section 5. Observationally, this dependence was first noted by Young & Knezek (1989), whose bins are also displayed in the figure. Their molecular fractions are generally higher, partly due to their rather high assumed X-factor of 2.8. The monotonic trend seems to break down between lenticular and elliptical galaxies, where the physical situation becomes more complex. In fact, many elliptical galaxies have molecular gas in their center with no detectable HI-counterpart, while others seem to have almost no H_2 (e.g. M 87, see Braine & Wiklind 1993), or may even exhibit HI-dominated outer regions left over by mergers (e.g. NGC 5266, see Morganti et al. 1997). To account for the different behavior of $R_{\rm mol}^{\rm galaxy}$ in elliptical and spiral galaxies, we chose a piecewise power-law with different powers for the two populations,

$$\log(R_{\rm mol,1}^{\rm galaxy}) = \left\{ \begin{array}{ll} c_1^{\rm el} + u_1^{\rm el}\,T & \text{if } T < T_1^* \\ c_1^{\rm sp} + u_1^{\rm sp}\,T & \text{if } T \geqslant T_1^* \end{array} \right\} + \sigma_{\rm phy,1} \tag{11}$$

where $c_1^{\rm el}$, $u_1^{\rm el}$, $c_1^{\rm sp}$, $u_1^{\rm sp}$ are considered as the free parameters to be fitted to the data, and T_1^* is at the intersection of the two regressions, i.e. $c_1^{\rm el} + u_1^{\rm el}\,T_1^* \equiv c_1^{\rm sp} + u_1^{\rm sp}\,T_1^*$, thus ensuring that $R_{\rm mol,1}^{\rm galaxy}$ remains a continuous function of T at $T = T_1^*$.

Another correlation exists between $R_{\rm mol}^{\rm galaxy}$ and the total cold gas mass $M_{\rm gas}$ or between $R_{\rm mol}^{\rm galaxy}$ and the blue magnitude $M_{\rm B}$. In fact, these two correlations are closely related due to the mutual correlation between $M_{\rm gas}$ and $M_{\rm B}$, and hence we shall restrict our considerations to the correlation between $R_{\rm mol}^{\rm galaxy}$ and $M_{\rm gas}$. According to the roughly monotonic trend visible in Fig. 5, we choose a power-law between $R_{\rm mol}^{\rm galaxy}$ and $M_{\rm gas}$ for our model 2,

$$\log(R_{\rm mol,2}^{\rm galaxy}) = q_2 + k_2\,\log\left(\frac{M_{\rm gas}}{10^9\,h^{-2}{\rm M}_\odot}\right) + \sigma_{\rm phy,2}, \tag{12}$$

where q_2, k_2 are free parameters. A somewhat similar dependence was recently found between $R_{\rm mol}^{\rm galaxy}$ and $M_{\rm HI}$ (Keres et al. 2003),

Model $\log(R^{\rm galaxy}_{\rm mol,i})$	$i = 0$	$i = 1$	$i = 2$	$i = 3$
$q_{\rm i}$	$-0.58^{+0.16}_{-0.23}$	-	$-0.51^{+0.03}_{-0.04}$	-
$c^{\rm el}_{\rm i}$	-	$+0.18^{+0.40}_{-0.22}$	-	$-0.01^{+0.25}_{-0.16}$
$u^{\rm el}_{\rm i}$	-	$+0.12^{+0.14}_{-0.05}$	-	$+0.13^{+0.07}_{-0.04}$
$c^{\rm sp}_{\rm i}$	-	$-0.14^{+0.10}_{-0.07}$	-	$-0.02^{+0.10}_{-0.09}$
$u^{\rm sp}_{\rm i}$	-	$-0.12^{+0.01}_{-0.02}$	-	$-0.13^{+0.02}_{-0.02}$
$k_{\rm i}$	-	-	$-0.24^{+0.05}_{-0.05}$	$-0.18^{+0.06}_{-0.07}$
$T^*_{\rm i}$	-	$-1.3^{+1.2}_{-0.5}$	-	$-0.1^{+1.2}_{-0.6}$
$\sigma_{\rm data,i}$	0.71	0.66	0.67	0.62
$\sigma_{\rm phy,i}$	0.39	0.27	0.30	0.15

Table 4. The upper panel lists the most likely parameters and 1-σ confidence intervals of the four models $R^{\rm galaxy}_{\rm mol,i}$ ($i = 0, ..., 3$). The bottom panel shows the rms-deviations $\sigma_{\rm data,i}$ of the data from the model predictions and the estimated physical scatter $\sigma_{\rm phy,i}$ for each model i.

but this result is less conclusive, since $R^{\rm galaxy}_{\rm mol}$ and $M_{\rm HI}$ are naturally correlated by the definition of $R^{\rm galaxy}_{\rm mol}$, even if $M_{\rm HI}$ and $M_{\rm H_2}$ are completely uncorrelated.

Finally, we shall introduce a fourth model (model 3) for $R^{\rm galaxy}_{\rm mol}$ that simultaneously depends on galaxy Hubble type and cold gas mass,

$$\log(R^{\rm galaxy}_{\rm mol,3}) = \left\{ \begin{array}{ll} c^{\rm el}_3 + u^{\rm el}_3\,T & (\text{if } T < T^*_3) \\ c^{\rm sp}_3 + u^{\rm sp}_3\,T & (\text{if } T \geqslant T^*_3) \end{array} \right\} + k_3 \log\left(\frac{M_{\rm gas}}{10^9\,h^{-2}{\rm M}_\odot}\right) + \sigma_{\rm phy,3}\,, \tag{13}$$

where $c^{\rm el}_3$, $u^{\rm el}_3$, $c^{\rm sp}_3$, $u^{\rm sp}_3$, k_3 are free parameters and T^*_3 is defined as $c^{\rm el}_3 + u^{\rm el}_3\,T^*_3 \equiv c^{\rm sp}_3 + u^{\rm sp}_3\,T^*_3$, thus making $R^{\rm galaxy}_{\rm mol,3}$ a continuous function of T at $T = T^*_3$. Comparing this model with models 1 and 2, will also allow us to study a possible degeneracy between model 1 and model 2 caused by a dependence between cold gas mass and galaxy Hubble type.

The free parameters of the above models were determined by minimizing the rms-deviation between the model predictions and the 245 observed values of $\log(R^{\rm galaxy}_{\rm mol})$ (Appendix A). Optimization in log-space is the most sensible choice since $R^{\rm galaxy}_{\rm mol}$ is subject to Gaussian scatter in log-space as will be shown in the Section 4.3. The most probable values of all parameters are shown in Table 4 together with the corresponding 1-σ confidence intervals. The latter were obtained using a bootstrapping method that uses 10^4 random half-sized subsamples of the full data set and determines the model-parameters for every one of them. The resulting distribution of values for each free parameter was approximated by a Gaussian distribution and its standard deviation σ was divided by $\sqrt{2}$ in order to find the 1-σ confidence intervals for the full data set. Note that in some cases the parameter uncertainties are coupled, i.e. a change in one parameter can be accommodated by changing the others, such that the model remains nearly identical. For models 1 and 2, the best fits are displayed in Figs. 4 and 5 as solid lines.

Table 4 also shows different scatters that will be explained in Section 4.3.

4.3 Scatter and uncertainty

The empirical values of $R^{\rm galaxy}_{\rm mol}$ scatter around the model predictions according to the distributions shown in Fig. 6 (dashed lines). The close similarity of these distributions to Gaussian distributions in log-space (solid lines) allows us to consider the rms-deviations of the data $\sigma_{\rm data}$ as the standard deviations of Gaussian distributions. This exhibits the advantage that $\sigma_{\rm data}$ can be decomposed in model-independent observational scatter $\sigma_{\rm obs}$ and model-dependent physical scatter $\sigma_{\rm phy}$ via the square-sum relation $\sigma^2_{\rm data,i} = \sigma^2_{\rm obs} + \sigma^2_{\rm phy,i}$, $i = 0, ..., 3$.

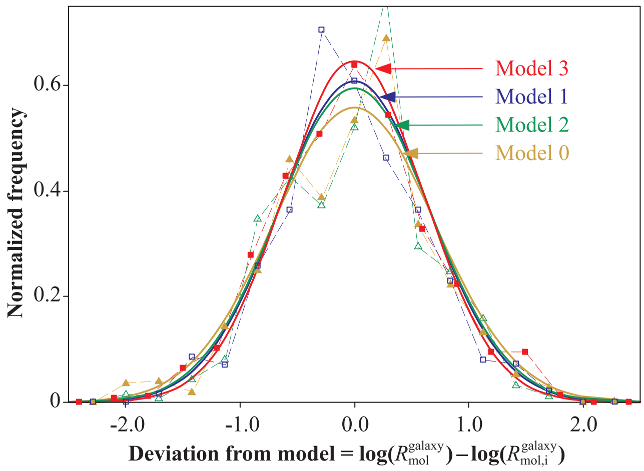

Figure 6. Distributions of the deviations between the observed values of $\log(R^{\rm galaxy}_{\rm mol})$ and the model-values $\log(R^{\rm galaxy}_{\rm mol,i})$ ($i = 1, ..., 3$). Data points and dashed lines represent the actual distribution of the data; solid lines represent Gaussian distributions with equal standard deviations.

The major contribution to $\sigma_{\rm data,i}$ comes from observational scatter, as suggested by the close similarity of the different values of $\sigma_{\rm data,i}$. Indeed, the observational scatter inferred from the $R^{\rm galaxy}_{\rm mol}$ values of the 22 repeated sources in our data is $\sigma_{\rm obs} \approx 0.6$. This scatter is a combination of CO-flux measurement uncertainties, uncertain CO/H_2-conversions and HI-flux uncertainties (in decreasing significance). Since $\sigma_{\rm obs}$ is only marginally smaller than $\sigma_{\rm data,i}$ for all models, the estimation of the physical scatters $\sigma_{\rm phy,i}$ (given in Table 4) is uncertain. Nevertheless, we shall include these best guesses of the physical scatter, when constructing the H_2-MFs in Section 4.4.

4.4 Recovering the H_2-MF and model evidence

Given a model for $R^{\rm galaxy}_{\rm mol}$, H_2-masses of arbitrary HI-galaxies can be estimated. We shall apply this technique to the 4315 sources in the HIPASS catalog using our four models of $R^{\rm galaxy}_{\rm mol,i}$, $i = 0, ..., 3$. For each model, the resulting H_2-catalog with 4315 objects will be converted into a H_2-MF, which can be compared to our reference H_2-MF derived directly from the CO-LF (Section 3).

For the models $R^{\rm galaxy}_{\rm mol,1}(T)$ and $R^{\rm galaxy}_{\rm mol,3}(M_{\rm gas}, T)$ Hubble types T were drawn from the HyperLeda database for each galaxy in the HIPASS catalog by means of the galaxy identifiers given in the optical cross-match catalog HOPCAT (Doyle et al. 2005). H_2-masses were then computed via $M_{\rm H_2} = R^{\rm galaxy}_{\rm mol,i}\,M_{\rm HI}$, $i = 0, ..., 3$. This equation is implicit in case of the mass-dependent models $R^{\rm galaxy}_{\rm mol,2}(M_{\rm gas})$ and $R^{\rm galaxy}_{\rm mol,3}(M_{\rm gas}, T)$, where $M_{\rm gas} = (M_{\rm HI} + M_{\rm H_2})/\beta$. All four models were applied with scatter, randomly drawn from a Gaussian distribution with the model-specific scatter $\sigma_{\rm phy,i}$, listed in Table 4.

In order to reconstruct a H_2-MF for each model, we employed the $1/V_{\rm max}$ method (Schmidt 1968), where $V_{\rm max}$ was calculated from the analytic completeness function for HIPASS that depends on the HI peak flux density $S_{\rm p}$, the integrated HI line flux $S_{\rm int}$, and the flux limit of the survey (Zwaan et al. 2004). After ensuring that

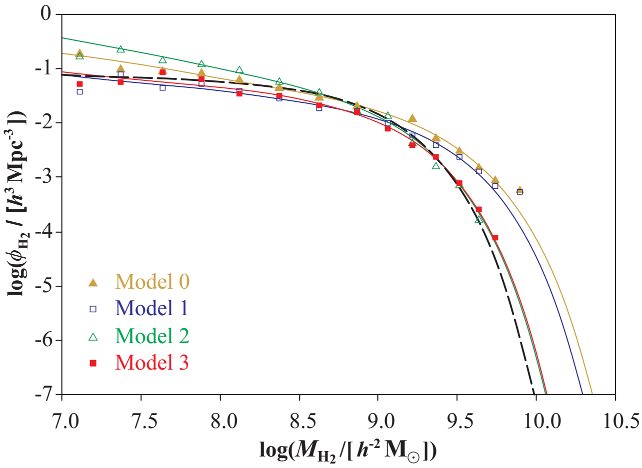

Figure 7. H_2-MFs constructed from the HIPASS HI-catalog using the different phenomenological models for the HI/H_2 ratio. The black dashed line is the reference H_2-MF derived from the CO-LF in Section 3.

we can accurately reproduce the HI-MF derived by Zwaan et al. (2005), we evaluated the four H_2-MFs (one for each model $R^{\rm galaxy}_{\rm mol,i}$) displayed in Fig. 7 (dots). The uncertainties of $\log(\phi_{H_2})$ vary around $\sigma = 0.03 - 0.1$. Each function was fitted by a Schechter function by minimizing the weighted rms-deviation (colored solid lines).

The comparison of these four H_2-MFs with the reference H_2-MF derived from the CO-LF allows us to qualify the different models $R^{\rm galaxy}_{\rm mol,i}$, $i = 0, ..., 3$, against each other. We ask: "What are the odds of model $R^{\rm galaxy}_{\rm mol,i}$ against model $R^{\rm galaxy}_{\rm mol,j}$ if the reference H_2-MF derived from the CO-LF is correct?" This question takes us back to the Bayesian framework of model selection applied in Section 2.2: If the models are a priori equally likely, their odds are equal to the Bayes factor, defined as the ratio between the model evidences. When computing these evidences, we take the "observational" data d to be the reference H_2-MF (with scatter), while the "model" data is the H_2-MF reproduced by applying a model $R^{\rm galaxy}_{\rm mol,i}$ to the HIPASS data. The free parameters θ (vector) are listed in Table 4 for each model (e.g. $c^{\rm el}_1$, $u^{\rm el}_1$, $c^{\rm sp}_1$, $u^{\rm sp}_1$ for model $R^{\rm galaxy}_{\rm mol,1}$). The prior probability density $\pi(\theta|M_{\rm i})$ in the evidence integral of Eq. (4) is taken as the multi-dimensional parameter probability distribution function obtained from the 245 galaxies studied in Section 4.2 (see Table 4). The second piece in the evidence integral, i.e. the probability density $p(d|\theta, M_{\rm i})$, is calculated as the product,

$$p(d|\theta, M_{\rm i}) = \prod_{\rm k} \frac{1}{\sigma\ \sqrt{2\pi}}\ \exp\left[\frac{(\phi^{\rm ref}_{\rm k} - \phi^{\rm model,i}_{\rm k})^2}{2\,\sigma^2}\right] \qquad (14)$$

where k labels the different bins of the H_2-MF (Fig. 7), and $\phi^{\rm ref}_{\rm k}$ and $\phi^{\rm model,i}_{\rm k}$ respectively denote the differential mass densities of the reference H_2-MF and the H_2-MFs reconstructed from HIPASS using the models $R^{\rm galaxy}_{\rm mol,i}$, $i = 0, ..., 3$. σ denotes the combined statistical uncertainties of $\phi^{\rm ref}_{\rm k}$ and $\phi^{\rm model,i}_{\rm k}$, theoretically given by $\sigma^2 = \sigma^{\rm ref\,2}_{\rm k} + \sigma^{\rm model,i\,2}_{\rm k}$. However, we shall neglect the contribution of $\sigma^{\rm model,i}_{\rm k}$, since $\sigma^{\rm ref}_{\rm k}$ is about $3-4$ times larger due to the small size of the FCRAO sample of CO-galaxies compared to the HIPASS sample of HI-galaxies. Furthermore, we assume that σ is independent of the bin k and adopt an average uncertainty equal to $\sigma = 0.15$ dex. This is the mean scatter of the binned data of the reference H_2-MF (see Fig. 3). Assuming a constant scatter for the whole reference MF artificially increases the weight of the low and high mass ends, where the scatter is indeed closer to 0.3 dex, and reduces the weight

Model	$R^{\rm galaxy}_{\rm mol,0}$	$R^{\rm galaxy}_{\rm mol,1}$	$R^{\rm galaxy}_{\rm mol,2}$	$R^{\rm galaxy}_{\rm mol,3}$
Nb. of free param.	1	4	2	5
$\ln B$	0.0	7.3	8.2	22

Table 5. Comparison of different models for H_2/HI-mass ratios of entire galaxies; the first row shows the number of free parameters, while the second row shows the model evidence in terms of the Bayes factor between that model and the constant model $R^{\rm galaxy}_{\rm mol,0}$.

of the central part, where the scatter equals 0.1 dex. We argue that this is a reasonable choice, since the central part of the reference H_2-MF suffers most from systematical uncertainties of the X-factor and the low and high mass ends encode much of the physics that could discriminate our models for $R^{\rm galaxy}_{\rm mol}$ against each other. In any case, the outcome of this evidence analysis is only weakly affected by the choice of scatter.

The integration of the evidence integral is computationally expensive: for *each* choice of model-parameters the following three steps need to be performed: (i) evaluation of the H_2-masses for each galaxies in the HIPASS sample, (ii) computation of the H_2-MF from that sample, (iii) computation of the product in Eq. (14). We applied a Monte Carlo method to sample the parameter spaces of the different models. About 10^6 integration steps had to be performed in total to reach a 2 per cent convergence of the Bayes factors.

The Bayes factor between each model $R^{\rm galaxy}_{\rm mol,i}$, $i = 0, ..., 3$, and $R^{\rm galaxy}_{\rm mol,0}$ is shown in Table 5: We find strong evidence for all variable models ($R^{\rm galaxy}_{\rm mol,1}$, $R^{\rm galaxy}_{\rm mol,2}$, $R^{\rm galaxy}_{\rm mol,3}$) against the constant one ($R^{\rm galaxy}_{\rm mol,0}$), and there is even stronger evidence of the bilinear model ($R^{\rm galaxy}_{\rm mol,3}$) against all others. The H_2-MF associated with this model is indeed the only one providing a simultaneous fit to the low and high mass ends of the reference MF (see Fig. 7), and the agreement is very good (reduced $\chi^2 = 0.8$).

On a physical level, there are good reasons for the partial failure of the other models in reproducing the extremities of the reference H_2-MF. Model $R^{\rm galaxy}_{\rm mol,1}(T)$ overestimates the space density of galaxies with high H_2-masses by overestimating $R^{\rm galaxy}_{\rm mol}$ for the gas-richest early-type spiral galaxies. In reality, the latter have a very low molecular fraction (see data, model $R^{\rm galaxy}_{\rm mol,2}$, theory in Section 5), but they are a minority within otherwise gas-poor but molecule-rich early-type spirals. Hence, a model depending on Hubble type alone is likely to miss out such objects, resulting in an increased density of high H_2-masses. While model $R^{\rm galaxy}_{\rm mol,2}(M_{\rm gas})$ overcomes this issue and produces the right density of high H_2-masses, it fails by a factor $3-4$ in the low-mass end ($M_{H_2} \lesssim 10^8 \mathrm{M}_\odot$). This is a direct manifestation of assigning high molecular fractions to all gas-poor galaxies, which neglects small young spirals with a dominant atomic phase. Finally, model $R^{\rm galaxy}_{\rm mol,0}$ seems to suffer from limitations at both ends of the H_2-MF.

The clear statistical evidence for model 3 shall be supported by the theoretical derivation of $R^{\rm galaxy}_{\rm mol}$ presented in Section 5.

5 THEORETICAL MODEL FOR THE H_2/HI-MASS RATIO

So far, we have approached the galactic H_2/HI-mass ratios $R^{\rm galaxy}_{\rm mol}$ with a set of *phenomenological* models, limited to the local Universe. By contrast, we have recently derived a *physical* model for the H_2/HI-ratios in regular galaxies, which potentially extends to high redshift (Obreschkow et al. 2009). This model relies on the theoretically and empirically established relation between interstellar gas pressure and local molecular fraction (Elmegreen 1993; Blitz & Rosolowsky 2006; Krumholz et al. 2009; Leroy et al. 2008). In this section, we will show that the physical model predicts H_2/HI-ratios consistent with our phenomenological model 3 given in Eq. (13). Hence, the physical (or "theoretical") model provides a reliable explanation for the global phenomenology of the H_2/HI-ratio in galaxies.

5.1 Background: the $R_{\rm mol}$–pressure relation

Understanding the observed continuous variation of $R_{\rm mol}$ within individual galaxies (e.g. Leroy et al. 2008) requires some explanation, since, fundamentally, there is no mixed thermodynamic equilibrium of HI and H_2. To first order, the ISM outside molecular clouds is atomic, while a cloud-region in local thermodynamic equilibrium (LTE) is either fully atomic or fully molecular, depending on the local state variables. The apparent continuous variation of $R_{\rm mol}$ is the combined result of (i) a non-resolved conglomeration of fully atomic and fully molecular clouds, (ii) clouds with molecular cores and atomic shells in different LTE, and (iii) some cloud regions off LTE with actual transient mixtures of HI and H_2. However, a time-dependent model for off-equilibrium clouds (Goldsmith et al. 2007) revealed that the characteristic time taken between the onset of cloud compression and full molecularization is of the order of 10^7 yrs, much smaller than the typical age of molecular clouds, and hence the fraction of these clouds is small. Therefore, averaged over galactic parts (hundreds or thousands of clouds), $R_{\rm mol}$ is dictated by clouds in LTE, entirely defined by a number of state variables.

A theoretical frame exploiting the LTE of molecular clouds was introduced by Elmegreen (1993), who considered an idealized double population of homogeneous diffuse clouds and isothermal self-gravitating clouds, both of which can have atomic and molecular shells. In this model the molecular mass fraction $f_{\rm mol} = {\rm d}M_{\rm H_2}/{\rm d}(M_{\rm HI} + M_{\rm H_2})$ of each cloud depends on the density profile and the photodissociative radiation density from stars j, corrected for self-shielding by the considered cloud, mutual shielding among different clouds, and dust extinction. Since the shielding from this radiation depends on the gas pressure, Elmegreen (1993) finds that $f_{\rm mol}$ essentially scales with the external pressure P and photodissociative radiation density j, approximately following $f_{\rm mol} \propto P^{2.2}\, j^{-1}$ with an asymptotic flattening towards $f_{\rm mol} = 1$ at high P and low j. This implies approximately $R_{\rm mol} \equiv {\rm d}M_{\rm H_2}/{\rm d}M_{\rm HI} \propto P^{2.2}\, j^{-1}$. Assuming that j is proportional to the surface density of stars $\Sigma_{\rm stars}$ and that the stellar velocity dispersion $\sigma_{\rm stars}$ varies radially as $\Sigma_{\rm stars}^{0.5}$, Wong & Blitz (2002) and Blitz & Rosolowsky (2004, 2006) find roughly $j \propto P$ and hence $R_{\rm mol} \propto P^{\alpha}$ with $\alpha = 1.2$. Recently, Krumholz et al. (2009) have presented a more elaborate theory concluding that $\alpha \approx 0.8$. However, the exponent α remains uncertain, thus requiring an empirical determination.

Observationally, Blitz & Rosolowsky (2004, 2006) were the first ones to reveal a surprisingly tight power-law relation between pressure and molecular fraction based on a sample of 14 nearby galaxies including dwarf galaxies, HI-rich galaxies, and H_2-rich galaxies. Perhaps the richest observational study published so far is the one by Leroy et al. (2008), who analyzed 23 galaxies of The HI Nearby Galaxy Survey (THINGS, Walter et al. 2008), for which H_2-densities had been derived from CO-data and star formation densities. This analysis confirmed the power-law relation

$$R_{\rm mol} = (P/P_*)^{\alpha}, \tag{15}$$

where P is the local, kinematic midplane pressure of the gas, and P_* and α are free parameters, whose best fit to the data is given by $P_* = 2.35 \cdot 10^{-13}$ Pa and $\alpha = 0.8$.

5.2 Physical model for the H_2/HI-ratio in galaxies

We shall now consider the consequence of the model given in Eq. (15) for the H_2/HI-ratio of entire galaxies. To this end, we adopt the models and methods presented in Obreschkow et al. (2009), restricting this paragraph to an overview.

First, we note that most cold gas of regular galaxies is normally contained in a disc. This even applies to bulge-dominated early-type galaxies, such as suggested by recently presented CO-maps of five nearby elliptical galaxies (Young 2002). Hence, the HI- and H_2-distributions of all regular galaxies can be well described by surface density profiles $\Sigma_{\rm HI}(r)$ and $\Sigma_{\rm H_2}(r)$. We assume that the disc is composed of axially symmetric, thin layers of stars and gas, which follow an exponential density profile with a generic scale length $r_{\rm disc}$, i.e.

$$\Sigma_{\rm stars}^{\rm disc}(r) \sim \Sigma_{\rm gas}(r) \sim \Sigma_{\rm HI}(r) + \Sigma_{\rm H_2}(r) \sim \exp(-r/r_{\rm disc}), \tag{16}$$

where r is the galactocentric radius and Σ denotes the mass column densities of the different components. Next, we adopt the phenomenological relation of Eq. (15), i.e.

$$\frac{\Sigma_{\rm HI}(r)}{\Sigma_{\rm H_2}(r)} = [P(r)/P_*]^{\alpha}, \tag{17}$$

and substitute the kinematic midplane pressure $P(r)$ for (Elmegreen 1989)

$$P(r) = \frac{\pi}{2}\, G\, \Sigma_{\rm gas}(r)\Big(\Sigma_{\rm gas}(r) + f\, \Sigma_{\rm stars}^{\rm disc}(r)\Big), \tag{18}$$

where G is the gravitational constant and $f \equiv \sigma_{\rm gas,z}/\sigma_{\rm stars,z}$ is the ratio between the vertical velocity dispersions of gas and stars. We adopt $f = 0.4$ according to Elmegreen (1989).

Eqs. (16, 17) can be solved for $\Sigma_{\rm HI}(r)$ and $\Sigma_{\rm H_2}(r)$. In Obreschkow et al. (2009), we demonstrate that the resulting surface profiles are consistent with the empirical data of the two nearby spiral galaxies NGC 5055 and NGC 5194 (Leroy et al. 2008). Integrating $\Sigma_{\rm HI}(r)$ and $\Sigma_{\rm H_2}(r)$ over the exponential disc gives the gas masses $M_{\rm HI}$ and $M_{\rm H_2}$, hence providing an estimate of their ratio $R_{\rm mol}^{\rm galaxy}$. Analytically, $R_{\rm mol}^{\rm galaxy}$ is given by an intricate expression, which is well approximated (relative error < 0.05 for all galaxies) by the double power-law

$$R_{\rm mol,th}^{\rm galaxy} = \left(3.44\, {R_{\rm mol}^{\rm c}}^{-0.506} + 4.82\, {R_{\rm mol}^{\rm c}}^{-1.054}\right)^{-1}, \tag{19}$$

where

$$R_{\rm mol}^{\rm c} = \left[11.3\,{\rm m}^4\,{\rm kg}^{-2}\,{\rm r}_{\rm disc}^{-4}\,{\rm M}_{\rm gas}\left({\rm M}_{\rm gas} + 0.4\,{\rm M}_{\rm stars}^{\rm disc}\right)\right]^{0.8}. \tag{20}$$

$R_{\rm mol}^{\rm c}$ is a dimensionless parameter, which can be interpreted as the H_2/HI-ratio at the center of a pure disc galaxy. For typical cold gas masses of average galaxies ($M_{\rm gas} = 10^8 - 10^{10} {\rm M}_\odot$) and corresponding stellar masses and scale radii, $R_{\rm mol}^{\rm c}$ calculated from Eq. (20) varies roughly between 0.1 and 50. Hence, $R_{\rm mol}^{\rm galaxy}$ given in Eq. (19) varies roughly between 0.01 and 1.

In summary, Eqs. (19, 20) represent a theoretical model for $R_{\rm mol}^{\rm galaxy}$, which uses three input parameters: the disc stellar mass $M_{\rm stars}^{\rm disc}$, the cold gas mass $M_{\rm gas}$, and the exponential scale radius $r_{\rm disc}$ (see Obreschkow et al. 2009 for a detailed discussion).

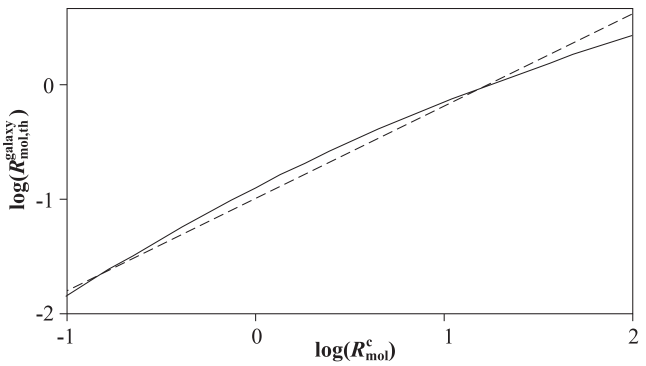

Figure 8. Visualization of the function $R^{\rm galaxy}_{\rm mol,th}(R^{\rm c}_{\rm mol})$. The solid line represents the nearly exact function given in Eq. (19), while the dashed line is the power-law fit of Eq. (21).

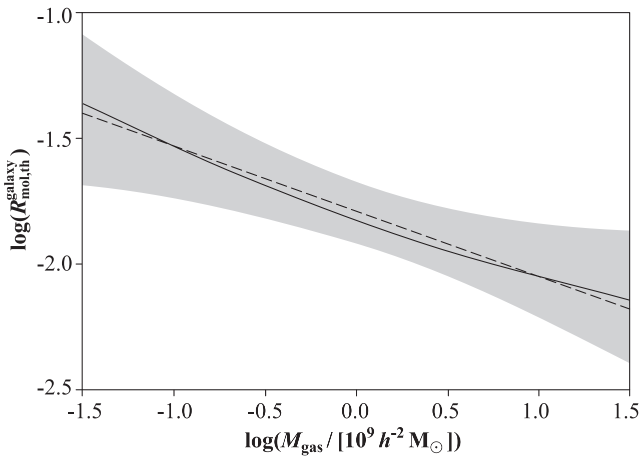

Figure 9. Relation between $R^{\rm galaxy}_{\rm mol,th}$ and $M_{\rm gas}$ for flat disks ($T = 10$). The solid line represents the relation obtained from Eq. (22), when expressing $r_{\rm disc}$ and $M^{\rm disc}_{\rm stars}$ as functions of $M_{\rm gas}$ using Eqs. (23, 24). The shaded zone represents the 1-σ uncertainty implied by the uncertainties of the empirical parameters in Eqs. (23, 24). The dashed line represents the best power-law fit for the displayed mass interval as given in Eq. (26).

5.3 Mapping between theoretical and phenomenological models

We shall now show that our theoretical model for galactic H_2/HI-mass ratios given in Eqs. (19, 20) closely matches the best phenomenological model given in Eq. (13). The mapping between the two models uses a list of empirical relations derived from observations of nearby spiral galaxies, and hence the comparison of the models is a priori restricted to spiral galaxies in the local Universe.

First, we note that Eq. (19) can be well approximated by the power-law

$$R^{\rm galaxy}_{\rm mol,th} \approx 0.1\,{R^{\rm c}_{\rm mol}}^{0.8}. \tag{21}$$

As shown in Fig. 8, this approximation is accurate to about 10 per cent over the whole range $R^{\rm c}_{\rm mol} = 0.1, ..., 50$, covering most regular galaxies in the local Universe.

Substituting $R^{\rm c}_{\rm mol}$ in Eq. (21) for Eq. (20), yields the approximate relation

$$R^{\rm galaxy}_{\rm mol,th} = \left[0.31\,{\rm m}^4\,{\rm kg}^{-2}\,{\rm r}^{-4}_{\rm disc}\,{\rm M}_{\rm gas}\left({\rm M}_{\rm gas} + 0.4\,{\rm M}^{\rm disc}_{\rm stars}\right)\right]^{0.64}. \tag{22}$$

In order to compare the theoretical model of $R^{\rm galaxy}_{\rm mol}$ to the empirical models of Section 4.2, we need to eliminate the formal dependence of $R^{\rm galaxy}_{\rm mol,th}$ on $r_{\rm disc}$ and $M^{\rm disc}_{\rm stars}$. To this end, we use two approximate empirical relations, derived from samples of nearby spiral galaxies (see Appendix B),

$$\log\left(\frac{M^{\rm disc}_{\rm stars}}{h^{-2}{\rm M}_\odot}\right) = \gamma_1 + \alpha_1\,\log\left(\frac{M_{\rm gas}}{2\cdot 10^9\,h^{-2}\,{\rm M}_\odot}\right), \tag{23}$$

$$\log\left(\frac{r_{\rm disc}}{h^{-1}{\rm kpc}}\right) = \gamma_2 + \alpha_2\,\log\left(\frac{M^{\rm disc}_{\rm stars}}{5\cdot 10^9\,h^{-2}\,{\rm M}_\odot}\right) + \delta\,\tilde{T}, \tag{24}$$

where $\tilde{T} \equiv (10-T)/16$ is the normalized Hubble type, which varies between $\tilde{T} = 0$ (pure disc galaxies) to $\tilde{T} = 1$ (pure spheroids).

The parameters corresponding to the best χ^2 fit (Appendix B) are $\alpha_1 = 1.46 \pm 0.1$, $\gamma_1 = 9.80 \pm 0.05$, $\alpha_2 = 0.45 \pm 0.05$, $\gamma_2 = 0.97 \pm 0.05$, $\delta = -1.07 \pm 0.1$. The given intervals are the 1-σ confidence intervals of the parameters; they do not characterize the scatter of the data. The units on the right hand side of Eqs. (23, 24) were chosen such as to minimize the correlations between the uncertainties of $\alpha_{\rm i}$ and $\gamma_{\rm i}$.

Physical reasons for the empirical relations in Eqs. (23, 24) are discussed in Appendix B. Substituting Eqs. (23, 24) into Eq. (22) reduces $R^{\rm galaxy}_{\rm mol,th}$ to a pure function of $M_{\rm gas}$ and T of the form

$$\log\left[R^{\rm galaxy}_{\rm mol,th}(M_{\rm gas}, T)\right] = \log\left[R^{\rm galaxy}_{\rm mol,th}(M_{\rm gas}, 10)\right] + \delta(0.16\,T - 1.6), \tag{25}$$

where $R^{\rm galaxy}_{\rm mol,th}(M_{\rm gas}, 10)$ is the theoretical H_2/HI-ratio of a pure disc galaxy, i.e. $T = 10$. The function $R^{\rm galaxy}_{\rm mol,th}(M_{\rm gas}, 10)$ is displayed in Fig. 9 together with the 1-σ uncertainty implied by the uncertainties of the four parameters α_1, α_2, γ_1, γ_2. We approximate this relation by the power-law

$$\log\left[R^{\rm galaxy}_{\rm mol,th}(M_{\rm gas}, 10)\right] = c + s\cdot\log\left(\frac{M_{\rm gas}}{10^9\,h^{-2}\,{\rm M}_\odot}\right). \tag{26}$$

The parameters minimizing the rms-deviation on the mass-interval $\log(M_{\rm gas}/[{\rm M}_\odot h^{-2}]) = 7.5 - 10.5$ are $c = -1.79 \pm 0.04$ and $s = -0.24 \pm 0.05$. The given uncertainties approximate the propagated uncertainties of α_1, α_2, γ_1, γ_2.

The simplified theoretical model for the H_2/HI-ratio given in Eqs. (25, 26) exhibits exactly the formal structure of our best phenomenological model 3. Setting $R^{\rm galaxy}_{\rm mol,th}(M_{\rm gas}, T)$ in Eq. (25) equal to $R^{\rm galaxy}_{\rm mol,3}(M_{\rm gas}, T)$ in Eq. (13) for spiral galaxies, yields the following mapping between the theoretical and empirical model-parameters,

$$\begin{aligned} c^{\rm sp}_3 &= c - 1.6\,\delta\,, \\ u^{\rm sp}_3 &= s\,, \\ k_3 &= 0.16\,\delta\,. \end{aligned} \tag{27}$$

The probability distributions of the empirical model-parameters on the left hand side of Eqs. (27) were derived in Section 4 and their 1-σ uncertainties are given in Table 4. The corresponding probability distributions of the theoretical model-parameters on the right hand side of Eqs. (27) can be estimated from the Gaussian uncertainties given for the parameters c, s, δ. The empirical and theoretical parameter distributions are compared in Fig. 10 and reveal a surprising consistency.

6 DISCUSSION

6.1 Theoretical versus phenomenological model

The dependence of $R^{\rm galaxy}_{\rm mol}$ on galaxy Hubble type T and cold gas mass $M_{\rm gas}$ was first considered on a purely phenomenological level,

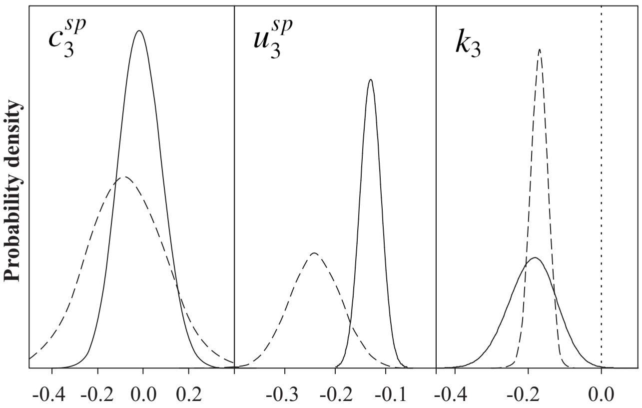

Figure 10. Probability distributions of the three parameters in our model 3 (Eq. 13) for the H_2/HI-mass ratio $R_{\rm mol}^{\rm galaxy}$ of spiral galaxies. Solid lines represent phenomenologically determined probability distributions given in Table 4; dashed lines represent the corresponding theoretical probability distributions, obtained when using Eqs. (27) with the respective distributions for c, s, and δ.

and described by the empirical models in Section 4. The best empirical model for spiral galaxies could be quantitatively reproduced by the subsequently derived theoretical model for regular galaxies in Section 5. Hence, the latter provides a tool for understanding the variations of $R_{\rm mol}^{\rm galaxy}$.

In fact, according to Eq. (22), $R_{\rm mol}^{\rm galaxy}$ seems most directly dictated by the scale radius $r_{\rm disc}$ and the masses $M_{\rm gas}$ and $M_{\rm stars}^{\rm disc}$. The dependence of $R_{\rm mol}^{\rm galaxy}$ on T is clearly due to the trend for smaller values of $r_{\rm disc}$ (for a given mass) in bulge-rich galaxies. Several physical reasons for the influence of the bulge on $r_{\rm disc}$ are mentioned in Appendix B2.

From Eq. (22), one might naively expect that $R_{\rm mol}^{\rm galaxy}$ and $M_{\rm gas}$ are positively correlated. However, the disc scale radius $r_{\rm disc}$ increases with $M_{\rm gas}$ as $r_{\rm disc} \propto M_{\rm gas}^{0.66}$ by virtue of Eqs. (23, 24). Taking this scaling into account, the H_2/HI-ratio $R_{\rm mol}^{\rm galaxy}$ effectively decreases with increasing $M_{\rm gas}$. The physical picture is that more massive galaxies are less dense due to their larger sizes, and hence their molecular fraction is lower.

The 'best' phenomenological model is by definition the one that, when applied to the galaxies in the HIPASS sample, exhibits the H_2-MF that best fits the reference H_2-MF derived from the CO-LF. The close agreement between the best model defined in this way and the theoretical model therefore supports the accuracy of the CO-LF (Keres et al. 2003), which could a priori be affected by the poorly characterized completeness of the CO-sample. Confirmingly, Keres et al. (2003) argued that the CO-LF does not substantially suffer from incompleteness by analyzing the FIR-LF produced from the same sample.

6.2 Brief word on cosmic evolution

The theoretical model $R_{\rm mol,th}^{\rm galaxy}$ given in Eqs. (19, 20) potentially extends to high redshift, as it only premises the invariance of the relation between pressure and $R_{\rm mol}$ and a few assumptions with weak dependence on redshift (but see discussion in Obreschkow et al. 2009). However, we emphasize that the transition from the theoretical model $R_{\rm mol,th}^{\rm galaxy}$ to the phenomenological model $R_{\rm mol,3}^{\rm galaxy}$ uses a set of relations extracted from observations in the local Universe. Most probably $R_{\rm mol,3}^{\rm galaxy}$ underestimates the molecular fraction at higher redshift, predominantly due to the evolution in the mass–diameter relation of Eq. (24). Indeed, scale radii are smaller at higher redshift for identical masses, thus increasing the pressure and molecular fraction. Bouwens et al. (2004) found $r_{\rm disc} \propto (1+z)^{-1}$ from observations in the Ultra Deep Field, consistent with the theoretical prediction by Mo et al. (1998). According to Eq. (22), where $R_{\rm mol}^{\rm galaxy} \propto r_{\rm disc}^{-2.6}$, this implies $R_{\rm mol}^{\rm galaxy} \propto (1+z)^{2.6}$. In other words, the phenomenological model 3 (Eq. 13) for the H_2/HI-mass ratio should be multiplied by roughly a factor $(1+z)^{2.6}$. However, this conclusion only applies if we consider galaxies with constant stellar and gas masses. For the cosmic evolution of the universal H_2/HI-ratio $R_{\rm mol}^{\rm universe}$, we also require a model for the evolution of the stellar and gas mass functions, and it may even be important to consider different scenarios for the evolution of the scale radius for different masses. A more elaborate model for the evolution of $R_{\rm mol}^{\rm universe}$ can be obtained from cosmological simulations (e.g. Obreschkow et al. 2009 and forthcoming publications).

6.3 Application: The local cold gas-MF

We finally apply our best phenomenological model for the H_2/HI-mass ratio (i.e. $R_{\rm mol,3}^{\rm galaxy}$ given in Eq. 13) to derive an integral cold gas-MF (HI+H_2+He) from the HIPASS catalog. In fact, the cold gas-MF cannot be inferred solely from the HI-MF (e.g. Zwaan et al. 2005) and the H_2-MF (e.g. Section 3), but only from a sample of galaxies with simultaneous HI- and H_2-data. Presently, there is no such sample with a large number of galaxies and an accurate completeness function. Therefore, we prefer using the HIPASS data, which have both sufficient size (4315 galaxies) and well described completeness (Zwaan et al. 2004), and we estimate the corresponding H_2-masses using our model $R_{\rm mol,3}^{\rm galaxy}$. Details of the computation of the H_2-masses were given in Section 4.4.

The resulting cold gas-MF is shown in Fig. 11 together with the HI-MF from Zwaan et al. (2005) and the reference H_2-MF derived in Section 3. The displayed continuous functions are best fitting Schechter functions. The respective Schechter function parameters for the cold gas-MF are $M^* = 7.21 \cdot 10^9\,h^{-2}\,{\rm M}_\odot$, $\alpha = -1.37$, and $\phi^* = 0.0114\,h^3\,{\rm Mpc}^{-3}$. The total cold gas density in the local Universe derived by integrating this Schechter function is $\Omega_{\rm gas} = 4.2 \cdot 10^{-4}\,h^{-1}$, closely matching the value $(4.4 \pm 0.8) \cdot 10^{-4}\,h^{-1}$ obtained when summing up the empirical HI-density (Zwaan et al. 2005), the H_2-density (Section 3), and the corresponding He-density.

7 CONCLUSION

In this paper, we established a coherent picture of the H_2/HI-ratio in galaxies based on a variety of extragalactic observations and theoretical considerations. Some important jigsaw pieces are:

(i) Measurements of the X-factor (summarized in Boselli et al. 2002) were combined with more recent CO-flux measurements and extinction-corrected optical $M_{\rm B}$-magnitudes, resulting in a working model for X.

(ii) This model for X was applied to the CO-LF by Keres et al. (2003) in order to derive the first local H_2-MF based on a variable X-factor.

(iii) Nine samples of local galaxies (245 objects in total) with simultaneous measurements of $M_{\rm HI}$ and $L_{\rm CO}$ were combined to fit a set of empirical models for galactic H_2/HI-mass ratios $R_{\rm mol}^{\rm galaxy}$.

(iv) These models were applied to the large HI-sample of the HIPASS catalog, which permitted the derivation of a H_2-MF for

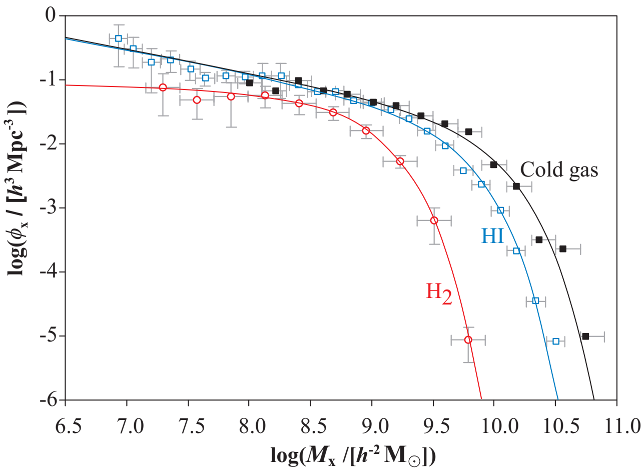

Figure 11. Filled squares represent the integral cold gas-MF (HI+H_2+He) derived from the HIPASS data using our best phenomenological model for the H_2/HI-mass ratio (Eq. 13); empty squares represent the observed HI-MF (Zwaan et al. 2005) and empty circles represent our best estimate of the H_2-MF (Section 3). Solid lines are best fitting Schechter functions.

each model for $R_{\rm mol}^{\rm galaxy}$. A comparison of these H_2-MFs with the one derived directly from the CO-LF allowed us to determine the statistical evidence of each model and to uncover a clear 'best model'.

(v) Based on the relation between pressure and the local H_2/HI-ratio $R_{\rm mol}$ (Leroy et al. 2008), we established a theoretical model for the H_2/HI-ratio $R_{\rm mol}^{\rm galaxy}$ of regular galaxies, which potentially extends to high redshifts.

(vi) We could show that the best empirical model for $R_{\rm mol}^{\rm galaxy}$ found before is an excellent approximation of the theoretical model in the local Universe.

The factual results standing out of this analysis are

(i) an empirical H_2-MF obtained by combining the CO-LF of Keres et al. (2003) with a variable X-factor (see Fig. 3 and parameters in Table 3),

(ii) an empirical model for $R_{\rm mol}^{\rm galaxy}$ (Eq. 13), which accurately reproduces the above H_2-MF, when applied to the HI-sample of the HIPASS catalog,

(iii) a theoretical model for $R_{\rm mol}^{\rm galaxy}$ (Eqs. 19, 20), which provides a source for physical understanding and generalizes to high redshift,

(iv) a quasi-empirical integral cold gas-MF (HI+H_2+He) based on the HIPASS data.

Self-consistency argues in favour of the interlinked picture established in this paper. However, all quantitative results remain subjected to the uncertainties of the X-factor. The latter appears as a scaling factor, affecting in the same way the reference H_2-MF derived from the CO-LF, the phenomenological models of $R_{\rm mol}^{\rm galaxy}$ and hence the H_2-MFs derived from HIPASS, as well as the P–$R_{\rm mol}$ relation and thus the theoretical model for $R_{\rm mol}^{\rm galaxy}$. In the future it may therefore be necessary to re-scale the quantitative results of this paper using a more accurate determination of X.

ACKNOWLEDGEMENTS

This effort/activity is supported by the European Community Framework Programme 6, Square Kilometre Array Design Studies (SKADS), contract no 011938. We further acknowledge the usage of the HyperLeda database (http://leda.univ-lyon1.fr) and we thank the anonymous referee for the helpful suggestions.

REFERENCES

Andreani P., Casoli F., Gerin M., 1995, A&A, 300, 43
Arimoto N., Sofue Y., Tsujimoto T., 1996, PASJ, 48, 275
Barnes D. G., et al., 2001, MNRAS, 322, 486
Bertsch D. L., Dame T. M., Fichtel C. E., Hunter S. D., Sreekumar P., Stacy J. G., Thaddeus P., 1993, ApJ, 416, 587
Blitz L., Fukui Y., Kawamura A., Leroy A., Mizuno N., Rosolowsky E., 2007, in Protostars and Planets V, Reipurth B., Jewitt D., Keil K., eds., pp. 81–96
Blitz L., Rosolowsky E., 2004, ApJL, 612, L29
—, 2006, ApJ, 650, 933
Boselli A., Lequeux J., Gavazzi G., 2002, Ap&SS, 281, 127
Bouwens R. J., Illingworth G. D., Blakeslee J. P., Broadhurst T. J., Franx M., 2004, ApJL, 611, L1
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Braine J., Wiklind T., 1993, A&A, 267, L47
Dalgarno A., 2000, in Molecular Hydrogen in Space, Combes F., Pineau Des Forets G., eds., p. 3
Davies J. I., 1990, MNRAS, 244, 8
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Doyle M. T., et al., 2005, MNRAS, 361, 34
Drapatz S., Zinnecker H., 1984, MNRAS, 210, 11P
Elmegreen B. G., 1989, ApJ, 338, 178
—, 1993, ApJ, 411, 170
Ferrière K. M., 2001, Reviews of Modern Physics, 73, 1031
Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518
Fukui Y., et al., 1999, in IAU Symp., Vol. 190, New Views of the Magellanic Clouds, Chu Y.-H., Suntzeff N., Hesser J., Bohlender D., eds., p. 61
Georgakakis A., Hopkins A. M., Caulton A., Wiklind T., Terlevich A. I., Forbes D. A., 2001, MNRAS, 326, 1431
Goldsmith P. F., Li D., Krčo M., 2007, ApJ, 654, 273
Guelin M., Zylka R., Mezger P. G., Haslam C. G. T., Kreysa E., Lemke R., Sievers A. W., 1993, A&A, 279, L37
Heyer M. H., Corbelli E., Schneider S. E., Young J. S., 2004, ApJ, 602, 723
Heyer M. H., Dame T. M., Thaddeus P., 2000, in Proc. 232. WE-Heraeus Seminar, Berkhuijsen E. M., Beck R., Walterbos R. A. M., eds., pp. 29–36
Hunter S. D., et al., 1997, ApJ, 481, 205
Israel F., 2000, in Molecular Hydrogen in Space, Combes F., Pineau Des Forets G., eds., p. 293
Israel F. P., 1997, A&A, 328, 471
Jeffreys H., 1961, Theory of Probability; 3 edition. Oxford University Press
Keres D., Yun M. S., Young J. S., 2003, ApJ, 582, 659
Kregel M., van der Kruit P. C., de Grijs R., 2002, MNRAS, 334, 646
Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216
Lauberts A., Valentijn E. A., 1989, The surface photometry catalogue of the ESO-Uppsala galaxies. Garching: European Southern Observatory
Ledoux C., Petitjean P., Srianand R., 2003, MNRAS, 346, 209
Lees J. F., Knapp G. R., Rupen M. P., Phillips T. G., 1991, ApJ, 379, 177
Leroy A., Bolatto A., Walter F., Blitz L., 2006, ApJ, 643, 825

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Matthews L. D., Gao Y., Uson J. M., Combes F., 2005, AJ, 129, 1849
McGaugh S. S., Bothun G. D., Schombert J. M., 1995, AJ, 110, 573
McGaugh S. S., de Blok W. J. G., 1997, ApJ, 481, 689
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Morganti R., Sadler E. M., Oosterloo T. A., Pizzella A., Bertola F., 1997, Publications of the Astronomical Society of Australia, 14, 89
Noterdaeme P., Ledoux C., Petitjean P., Srianand R., 2008, A&A, 481, 327
Obreschkow D., Croton D., DeLucia G., Khochfar S., Rawlings S., 2009, ApJ, 698, 1467
Paturel G., Petit C., Prugniel P., Theureau G., Rousseau J., Brouty M., Dubois P., Cambrésy L., 2003, A&A, 412, 45
Rubio M., Garay G., Montani J., Thaddeus P., 1991, ApJ, 368, 173
Sage L. J., 1993, A&A, 272, 123
Sage L. J., Welch G. A., 2006, ApJ, 644, 850
Sauty S., et al., 2003, A&A, 411, 381
Schechter P., 1976, ApJ, 203, 297
Schmidt M., 1968, ApJ, 151, 393
Sivia D., Skilling J., 2006, Data Analysis: A Bayesian Tutorial; 2 edition. Oxford University Press
Thronson H. A. J., Tacconi L., Kenney J., Greenhouse M. A., Margulis M., Tacconi-Garman L., Young J. S., 1989, ApJ, 344, 747
Valentijn E. A., van der Werf P. P., 1999, ApJL, 522, L29
Verheijen M., van Gorkom J. H., Szomoru A., Dwarakanath K. S., Poggianti B. M., Schiminovich D., 2007, ApJL, 668, L9
Wall W. F., 2006, Rev. Mex. Astron. Astrofis., 42, 117
Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Thornley M. D., Leroy A., 2008, AJ, 136, 2563
Welch G. A., Sage L. J., 2003, ApJ, 584, 260
Wild W., Harris A. I., Eckart A., Genzel R., Graf U. U., Jackson J. M., Russell A. P. G., Stutzki J., 1992, A&A, 265, 447
Wong T., Blitz L., 2002, ApJ, 569, 157
Young J. S., Knezek P. M., 1989, ApJL, 347, L55
Young J. S., Scoville N. Z., 1991, ARA&A, 29, 581
Young J. S., Xie S., Kenney J. D. P., Rice W. L., 1989, ApJS, 70, 699
Young J. S., et al., 1995, ApJS, 98, 219
Young L. M., 2002, AJ, 124, 788
Yun M. S., Reddy N. A., Condon J. J., 2001, ApJ, 554, 803
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30
Zwaan M. A., et al., 2004, MNRAS, 350, 1210

APPENDIX A: HOMOGENIZED DATA

This section presents the data (245 galaxies) used for the derivation of the models of $R_{\rm mol}^{\rm galaxy}$ in section 4.

CO-luminosities were drawn from 10 smaller samples: 17 nearby ($\lesssim$ 10 Mpc) lenticulars and ellipticals (Welch & Sage 2003; Sage & Welch 2006), 4 late-type spirals (Matthews et al. 2005), 68 isolated late-type spirals (Sauty et al. 2003), 6 ellipticals (Georgakakis et al. 2001), 17 spirals of all types (Andreani et al. 1995), 48 nearby ($\lesssim$ 10 Mpc) spirals of all types (Sage 1993), 12 ellipticals (Lees et al. 1991), 18 lenticulars and ellipticals (Thronson et al. 1989), 77 spirals of all types (Young & Knezek 1989). These 267 objects contained 22 repeated galaxies. In each case of repetition, the older reference was removed, such as to remain with the 245 distinct sources listed in Table A1. The CO-luminosities were homogenized by making them independent of different X-factors and Hubble constants. All other properties listed in the table were taken from homogenized reference catalogs, such as described in Section 4.1.

Table A1: Homogenized galaxy sample based on data drawn from the literature. T is the numerical Hubble type (see online help of the HyperLeda database), D_l the luminosity distance, M_B is the extinction corrected absolute blue magnitude, and X is the variable X-factor derived from M_B (Eq. 8) without addition of Gaussian scatter. The references for H_2-masses are: [1] Welch & Sage (2003); Sage & Welch (2006), [2] Matthews et al. (2005), [3] Sauty et al. (2003), [4] Georgakakis et al. (2001), [5] Andreani et al. (1995), [6] Sage (1993), [7] Lees et al. (1991), [8] Thronson et al. (1989), [9] Young & Knezek (1989).

Object	T	$D_l\,/\,h^{-1}$ Mpc	$M_B - 5\log h$	X	$\log(M_{H_2}/X\,h^{-2}\,M_\odot)$	Ref. H_2	$\log(M_{HI}/h^{-2}\,M_\odot)$
NGC 404	-2.8	1.7	-15.86	6.66	6.06	1	7.51
NGC 2787	-1.1	9.5	-18.87	2.21	6.58	1	8.58
NGC 3115	-2.8	6.4	-19.27	1.91	5.60	1	6.75
NGC 3384	-2.7	9.2	-19.06	2.06	5.87	1	5.94
NGC 3489	-1.3	7.7	-18.45	2.58	6.12	1	6.46
NGC 3607	-3.1	10.3	-19.23	1.94	8.34	1	6.93
NGC 3870	-2.0	9.9	-16.56	5.15	7.44	1	8.08
NGC 3941	-2.0	11.0	-19.04	2.08	7.15	1	8.81
NGC 4026	-1.8	12.1	-18.82	2.25	7.27	1	7.86
NGC 4150	-2.1	6.8	-17.66	3.44	6.91	1	6.88
NGC 4203	-2.7	12.7	-18.86	2.22	6.21	1	8.41
NGC 4310	-1.0	10.8	-16.86	4.61	6.96	1	7.10
NGC 4460	-0.9	7.3	-17.04	4.32	6.45	1	8.26
NGC 4880	-1.5	14.8	-17.92	3.13	6.27	1	6.02
NGC 7013	0.5	9.6	-18.79	2.28	7.30	1	8.70
NGC 7077	-3.9	12.0	-16.13	6.03	6.09	1	7.60
NGC 7457	-2.6	9.6	-18.29	2.73	5.85	1	5.88
NGC 100	5.9	9.0	-17.61	3.51	5.91	2	8.87
UGC 2082	5.9	7.7	-17.72	3.37	5.89	2	8.80
UGC 3137	4.2	12.5	-17.05	4.30	6.20	2	9.11
UGC 6667	6.0	12.1	-17.06	4.29	5.73	2	8.54
UGC 5	3.9	74.4	-20.98	1.02	8.76	3	9.82
NGC 7817	4.1	24.1	-20.42	1.25	8.44	3	9.30
IC 1551	3.6	136.0	-22.17	0.66	8.94	3	9.34
NGC 237	4.5	42.0	-19.81	1.57	8.53	3	9.75
NGC 575	5.3	32.3	-19.10	2.03	8.04	3	9.18
NGC 622	3.4	52.1	-19.93	1.50	8.24	3	9.54
UGC 1167	5.9	43.6	-19.18	1.97	8.85	3	9.61
UGC 1395	3.1	52.3	-19.90	1.51	8.43	3	9.25
UGC 1587	3.7	57.4	-20.38	1.27	7.86	3	9.59
UGC 1706	5.8	49.4	-19.82	1.56	7.96	3	9.17
IC 302	4.1	59.6	-21.33	0.90	8.43	3	10.19
IC 391	4.9	18.3	-18.91	2.18	7.46	3	8.89
UGC 3420	3.1	54.5	-20.96	1.03	8.03	3	10.01
UGC 3581	5.2	53.2	-20.30	1.31	8.24	3	9.56
NGC 2344	4.4	11.3	-17.91	3.14	6.73	3	8.66
UGC 3863	1.1	62.2	-20.53	1.20	8.32	3	9.30
UGC 4684	7.2	24.9	-17.92	3.13	6.82	3	9.11
NGC 2746	1.1	73.7	-20.65	1.15	8.65	3	9.64
UGC 4781	5.9	14.4	-16.54	5.19	6.46	3	8.90
UGC 5055	3.1	79.4	-20.19	1.36	8.79	3	10.02
NGC 2900	5.9	54.3	-19.51	1.75	8.57	3	9.69
NGC 2977	3.2	33.5	-19.95	1.49	8.31	3	8.83
NGC 3049	2.5	15.0	-17.86	3.20	7.24	3	8.86
IC 651	8.2	45.2	-20.37	1.28	8.54	3	9.53
NGC 3526	5.2	14.5	-18.68	2.37	7.73	3	8.64
UGC 6568	8.2	60.8	-19.86	1.54	8.12	3	9.14
UGC 6769	3.0	88.2	-20.66	1.15	9.10	3	9.96
UGC 6780	6.4	17.3	-16.79	4.73	7.29	3	9.28
UGC 6879	7.1	24.1	-18.78	2.28	7.82	3	8.83
UGC 6903	5.9	19.3	-17.69	3.40	7.46	3	9.07

Object	T	$D_l\,/[\,h^{-1}\,\mathrm{Mpc}]$	$M_B - 5\log h$	X	$\log(M_{H_2}/[X\,h^{-2}\,\mathrm{M}_\odot])$	Ref. H_2	$\log(M_{HI}/[h^{-2}\,\mathrm{M}_\odot])$
NGC 4348	4.1	20.3	-19.49	1.76	8.10	3	9.01
NGC 4617	3.1	49.6	-20.70	1.13	8.56	3	9.90
NGC 4635	6.5	10.9	-17.28	3.96	6.73	3	8.23
NGC 5377	1.1	20.6	-19.83	1.55	7.81	3	8.91
NGC 5375	2.4	26.0	-19.54	1.73	7.60	3	9.24
NGC 5584	5.9	17.1	-19.06	2.06	7.22	3	9.27
NGC 5690	5.4	18.4	-19.88	1.53	8.15	3	9.33
NGC 5768	5.3	20.3	-18.74	2.32	7.90	3	9.11
NGC 5772	3.1	52.3	-20.41	1.26	8.25	3	9.49
NGC 5913	1.3	20.8	-19.00	2.11	8.22	3	8.44
NGC 6012	1.9	20.1	-19.00	2.11	7.73	3	9.26
IC 1231	5.8	55.9	-20.71	1.13	8.04	3	9.14
UGC 10699	4.4	65.5	-20.19	1.36	8.60	3	9.11
UGC 10743	1.1	27.2	-18.75	2.31	7.52	3	8.78
NGC 6347	3.1	64.3	-20.46	1.23	8.57	3	9.48
UGC 10862	5.3	18.2	-17.81	3.26	7.21	3	9.07
NGC 6389	3.6	33.1	-20.37	1.28	8.30	3	9.93
UGC 11058	3.2	50.6	-20.48	1.22	8.51	3	9.40
NGC 6643	5.2	17.8	-20.31	1.30	8.35	3	9.27
NGC 6711	4.0	50.1	-20.18	1.37	8.77	3	9.14
UGC 11635	3.7	51.8	-21.05	0.99	8.95	3	9.88
UGC 11723	3.1	50.1	-19.87	1.53	8.48	3	9.57
NGC 7056	3.6	55.8	-20.53	1.20	8.67	3	9.11
NGC 7156	5.9	40.8	-20.12	1.40	8.43	3	9.32
UGC 11871	3.1	82.9	-20.38	1.27	9.22	3	9.43
NGC 7328	2.1	29.2	-19.31	1.88	8.34	3	9.45
NGC 7428	1.1	31.0	-18.85	2.23	7.72	3	9.44
UGC 12304	5.2	35.3	-19.40	1.82	8.01	3	8.88
UGC 12372	4.0	57.7	-19.94	1.49	8.65	3	9.49
NGC 7514	4.3	51.1	-20.62	1.16	8.18	3	9.16
UGC 12474	1.1	53.5	-20.53	1.20	8.80	3	8.87
NGC 7664	5.1	36.3	-20.03	1.44	8.51	3	9.91
UGC 12646	3.0	83.7	-20.84	1.07	8.68	3	9.70
NGC 7712	1.6	31.9	-18.94	2.15	7.84	3	9.10
IC 1508	7.2	43.8	-20.07	1.42	8.45	3	9.75
UGC 12776	3.0	51.8	-19.88	1.53	8.31	3	9.99
IC 5355	5.7	50.8	-19.56	1.72	8.26	3	9.05
UGC 12840	-1.8	71.3	-20.27	1.32	7.97	3	9.43
NGC 2623	2.0	57.2	-20.59	1.18	9.02	4	9.01
NGC 2865	-4.1	26.0	-20.01	1.46	7.35	4	8.79
NGC 3921	0.0	61.9	-21.00	1.01	8.82	4	9.46
NGC 4649	-4.6	12.1	-20.70	1.13	7.15	4	8.35
NGC 7252	-2.1	47.0	-20.73	1.12	8.83	4	9.29
NGC 7727	1.1	17.9	-19.98	1.47	7.27	4	8.45
NGC 142	3.1	81.4	-20.46	1.23	9.36	5	9.43
IC 1553	5.4	28.0	-18.75	2.31	7.69	5	9.10
ESO 473-27	4.4	193.6	-21.12	0.97	9.78	5	9.75
NGC 232	1.1	66.7	-19.82	1.56	9.50	5	9.21
ESO 475-16	2.1	70.7	-20.44	1.24	9.01	5	9.74
NGC 578	5.0	14.8	-19.73	1.61	7.97	5	9.52
ESO 478-6	4.1	52.6	-20.64	1.16	8.96	5	9.23
NGC 1187	5.0	12.2	-19.39	1.83	8.69	5	9.33
NGC 1306	2.8	12.7	-16.85	4.63	7.34	5	8.64
NGC 1385	5.9	13.1	-19.56	1.72	8.59	5	9.07
ESO 549-23	1.2	40.8	-19.42	1.81	8.45	5	8.88
ESO 483-12	0.3	41.0	-19.18	1.97	8.27	5	8.83
NGC 1591	1.9	39.5	-19.65	1.66	8.51	5	9.04
NGC 7115	3.4	34.1	-19.53	1.73	8.26	5	9.52
NGC 7225	-0.5	47.9	-20.09	1.41	9.29	5	9.07
NGC 7314	4.0	13.2	-19.71	1.62	8.05	5	9.24

Object	T	D_l /[h^{-1} Mpc]	$M_B - 5 \log h$	X	$\log(M_{H_2}/[X\,h^{-2}\,M_\odot])$	Ref. H_2	$\log(M_{HI}/[h^{-2}\,M_\odot])$
NGC 628	5.2	6.9	-19.84	1.55	8.55	6	9.73
NGC 672	6.0	5.1	-19.03	2.08	6.60	6	9.07
NGC 891	3.0	6.7	-19.43	1.80	8.97	6	9.72
NGC 925	7.0	6.6	-19.32	1.87	8.04	6	9.57
NGC 1058	5.3	6.3	-17.78	3.29	7.42	6	8.93
NGC 1560	7.0	2.3	-15.91	6.53	5.88	6	8.47
NGC 2403	6.0	3.2	-18.89	2.19	7.31	6	9.54
NGC 2683	3.1	5.2	-19.53	1.73	7.63	6	8.54
NGC 2903	4.0	6.3	-20.16	1.38	8.39	6	9.01
NGC 2976	5.3	1.6	-17.35	3.86	6.42	6	7.49
NGC 3031	2.4	2.4	-19.90	1.51	7.42	6	9.15
NGC 3184	5.9	7.7	-19.11	2.02	8.35	6	9.11
NGC 3344	4.0	6.9	-18.89	2.19	7.74	6	9.01
NGC 3351	3.0	8.3	-19.46	1.78	8.08	6	8.67
NGC 3368	1.8	9.4	-20.12	1.40	8.18	6	8.95
NGC 3486	5.2	8.2	-18.84	2.23	7.50	6	9.03
NGC 3521	4.0	8.0	-20.31	1.30	8.75	6	9.63
NGC 3593	-0.4	6.9	-17.50	3.65	7.62	6	7.75
NGC 3623	1.0	8.9	-20.17	1.37	7.62	6	8.27
NGC 3627	3.0	7.9	-20.40	1.26	8.55	6	8.56
NGC 3628	3.1	9.2	-20.67	1.14	8.62	6	9.33
NGC 4020	6.9	9.2	-17.31	3.91	6.60	6	8.05
NGC 4062	5.3	9.4	-18.78	2.28	7.63	6	8.47
NGC 4096	5.3	7.9	-19.49	1.76	7.75	6	8.86
NGC 4144	6.0	3.1	-15.93	6.48	6.31	6	8.09
NGC 4244	6.1	2.3	-18.06	2.97	6.62	6	8.72
NGC 4245	0.1	10.5	-17.97	3.07	7.39	6	6.61
NGC 4274	1.7	10.9	-19.33	1.87	8.27	6	8.75
NGC 4288	7.1	7.5	-16.32	5.62	6.67	6	8.52
NGC 4314	1.0	11.5	-19.02	2.09	7.69	6	6.43
NGC 4359	5.0	14.3	-17.49	3.66	6.55	6	8.44
NGC 4414	5.1	8.9	-19.25	1.92	8.48	6	8.90
NGC 4448	1.8	8.2	-17.86	3.20	7.39	6	7.38
NGC 4490	7.0	8.0	-20.93	1.04	7.45	6	9.54
NGC 4437	6.0	11.6	-20.70	1.13	8.14	6	7.90
NGC 4525	5.9	13.5	-18.11	2.92	6.57	6	7.86
NGC 4559	6.0	9.8	-20.35	1.28	8.26	6	9.57
NGC 4565	3.2	13.8	-21.74	0.77	8.62	6	9.48
NGC 4605	4.9	3.0	-17.58	3.54	6.82	6	8.05
NGC 4631	6.6	7.9	-21.46	0.86	8.03	6	9.58
NGC 4736	2.4	5.3	-19.27	1.91	7.86	6	8.23
NGC 4826	2.4	5.5	-19.86	1.54	7.79	6	8.07
NGC 5055	4.0	7.3	-20.43	1.25	8.80	6	9.40
NGC 5194	4.0	7.0	-19.74	1.61	9.29	6	9.21
NGC 5457	5.9	5.0	-20.26	1.33	8.50	6	9.79
NGC 6503	5.9	4.6	-17.77	3.31	7.35	6	8.86
NGC 6946	5.9	4.1	-20.12	1.40	8.74	6	9.55
NGC 7640	5.3	5.5	-18.75	2.31	6.93	6	9.62
NGC 185	-4.8	0.7	-13.83	14.00	4.81	7	5.18
NGC 205	-4.7	0.7	-13.61	15.18	4.95	7	5.57
NGC 855	-4.6	6.9	-16.23	5.81	5.33	7	7.62
NGC 3265	-4.8	15.7	-17.28	3.96	7.13	7	7.95
NGC 3928	-4.5	12.1	-17.35	3.86	7.36	7	8.22
NGC 5128	-2.1	5.3	-20.59	1.17	8.16	7	8.28
NGC 5666	6.4	23.6	-18.90	2.19	8.00	7	8.63
NGC 1819	-1.9	44.8	-20.23	1.34	9.10	7	9.13
NGC 3032	-1.8	16.7	-18.14	2.89	7.72	7	7.76
NGC 4138	-0.9	10.9	-17.97	3.07	7.13	7	8.54
NGC 7465	-1.9	20.6	-18.57	2.47	8.11	7	9.20
NGC 3413	-1.8	7.9	-16.66	4.96	7.21	8	7.95

Object	T	$D_l\,/[\,h^{-1}\,\mathrm{Mpc}]$	$M_\mathrm{B} - 5\,\log\,h$	X	$\log(M_{\mathrm{H}_2}/[X\,h^{-2}\,\mathrm{M}_\odot])$	Ref. H_2	$\log(M_\mathrm{HI}/[h^{-2}\,\mathrm{M}_\odot])$
NGC 5866	-1.2	9.5	-19.23	1.94	7.81	8	8.15
NGC 4710	-0.8	13.8	-19.02	2.09	8.25	8	7.20
NGC 4459	-1.4	13.3	-19.37	1.84	8.30	8	6.70
NGC 4526	-1.9	6.7	-18.63	2.41	8.30	8	7.05
NGC 693	0.1	15.5	-18.08	2.95	7.53	8	8.74
NGC 2685	-1.1	11.0	-18.32	2.70	7.45	8	8.79
NGC 2273	1.0	20.7	-19.47	1.77	8.26	8	8.90
NGC 3611	1.1	16.1	-18.67	2.38	8.42	8	8.75
NGC 4457	0.4	9.4	-18.31	2.71	8.63	8	8.27
NGC 4383	1.0	18.3	-19.01	2.10	7.91	8	9.15
NGC 7625	1.2	17.2	-18.38	2.64	8.56	8	8.98
NGC 23	1.2	47.4	-20.84	1.07	9.30	9	9.69
NGC 253	5.1	1.7	-20.19	1.36	8.32	9	9.04
NGC 520	0.8	21.5	-19.90	1.51	9.35	9	9.50
NGC 828	1.0	55.9	-20.95	1.03	9.75	9	9.80
NGC 834	3.9	48.1	-20.32	1.30	9.13	9	9.47
NGC 864	5.1	15.4	-19.80	1.57	8.49	9	9.78
NGC 877	4.8	39.7	-21.15	0.96	9.34	9	10.08
NGC 1055	3.2	9.3	-18.97	2.13	9.37	9	9.39
IC 342	5.9	2.3	-19.85	1.54	8.70	9	9.68
NGC 1530	3.1	27.5	-20.70	1.13	9.10	9	9.76
NGC 1569	9.6	2.4	-15.94	6.46	5.89	9	8.09
NGC 1614	4.9	47.2	-20.64	1.16	9.36	9	9.28
NGC 2146	2.3	11.6	-20.34	1.29	9.04	9	9.50
NGC 2339	4.0	22.9	-20.02	1.45	9.27	9	9.45
NGC 2276	5.4	27.2	-20.80	1.09	9.31	9	9.50
NGC 2532	5.2	54.5	-21.00	1.01	9.10	9	9.92
NGC 2633	3.0	24.5	-19.47	1.77	8.83	9	9.41
NGC 2775	1.7	13.5	-19.83	1.55	8.30	9	8.16
NGC 2841	3.0	8.3	-20.07	1.42	8.61	9	9.20
NGC 3034	8.0	1.7	-17.30	3.93	8.16	9	8.54
NGC 3079	6.6	13.5	-20.68	1.14	9.16	9	9.57
NGC 3147	3.9	31.1	-21.43	0.86	9.65	9	9.52
NGC 3221	5.6	42.5	-19.86	1.54	9.24	9	9.81
NGC 3310	4.0	12.2	-19.26	1.92	7.81	9	9.33
NGC 3437	5.2	13.8	-19.03	2.08	7.91	9	9.03
NGC 3504	2.1	16.8	-19.68	1.64	8.50	9	8.37
NGC 3556	6.0	9.3	-19.89	1.52	8.37	9	9.35
NGC 3893	5.1	12.0	-20.13	1.39	8.35	9	9.29
NGC 4192	2.5	10.0	-20.83	1.08	8.57	9	9.33
NGC 4194	9.7	27.3	-19.87	1.53	8.61	9	8.96
NGC 4254	5.2	25.2	-21.82	0.75	9.07	9	9.39
NGC 4303	4.0	16.2	-21.05	0.99	8.95	9	9.38
NGC 4321	4.0	16.8	-21.29	0.91	9.12	9	9.06
NGC 4388	2.8	26.2	-21.16	0.95	7.96	9	8.33
NGC 4394	2.9	10.3	-18.65	2.39	8.04	9	8.22
NGC 4402	3.3	10.0	-17.18	4.09	8.39	9	8.23
NGC 4419	1.1	10.0	-18.43	2.60	8.56	9	7.62
NGC 4424	1.2	5.1	-15.75	6.91	7.34	9	7.89
NGC 4438	0.7	10.0	-19.99	1.47	7.92	9	8.26
NGC 4449	9.8	2.6	-16.83	4.66	6.60	9	9.10
NGC 4450	2.3	20.7	-21.10	0.98	8.25	9	7.95
NGC 4501	3.4	23.9	-22.33	0.62	8.94	9	8.91
NGC 4527	4.0	17.9	-20.75	1.11	8.85	9	9.35
NGC 4535	5.0	20.3	-21.18	0.95	8.79	9	9.41
NGC 4536	4.2	18.6	-21.02	1.01	8.46	9	9.23
NGC 4548	3.1	6.0	-20.01	1.46	8.33	9	8.68
NGC 4569	2.4	10.0	-20.33	1.29	8.77	9	8.32
NGC 4571	6.3	10.0	-17.53	3.60	8.17	9	8.49
NGC 4579	2.8	16.1	-20.91	1.05	8.55	9	8.38

Object	T	$D_l\,/[\,h^{-1}\,\mathrm{Mpc}]$	$M_B - 5\,\log\,h$	X	$\log(M_{H_2}/[X\,h^{-2}\,M_\odot])$	Ref. H_2	$\log(M_{HI}/[h^{-2}\,M_\odot])$
NGC 4647	5.2	14.9	-19.02	2.09	8.37	9	8.33
NGC 4651	5.2	9.1	-18.86	2.22	8.14	9	9.21
NGC 4654	5.9	11.4	-19.86	1.54	8.46	9	9.15
NGC 4689	4.7	17.2	-19.92	1.50	8.44	9	8.30
NGC 5236	5.0	4.5	-19.99	1.46	9.41	9	9.86
NGC 5936	3.2	42.2	-20.29	1.31	9.15	9	8.90
NGC 6207	4.9	10.9	-19.11	2.02	7.52	9	8.97
NGC 6574	3.9	24.6	-20.12	1.40	9.13	9	8.87
NGC 7217	2.5	11.2	-19.70	1.63	8.41	9	8.65
NGC 7331	3.9	9.9	-20.81	1.09	9.22	9	9.77
NGC 7469	1.1	50.4	-21.00	1.01	9.50	9	9.30
NGC 7479	4.4	24.6	-20.93	1.04	9.35	9	9.77
NGC 7541	4.7	27.2	-20.78	1.10	9.34	9	10.01
NGC 7674	3.8	91.6	-21.17	0.95	9.66	9	10.11

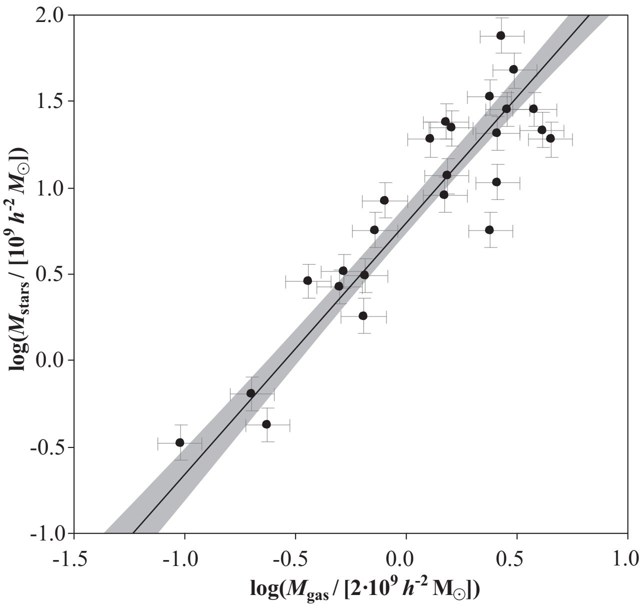

Figure B1. Data points (subsample of the data shown in Appendix A) represent the observed relation between disc stellar mass $M_{\rm stars}^{\rm disc}$ and cold gas mass $M_{\rm gas}$. The solid line shows the best power-law fit and the shaded envelope its 1-σ uncertainty. This power-law is given in Eq. (23) and has a slope of $\alpha_1 = 1.46 \pm 0.10$.

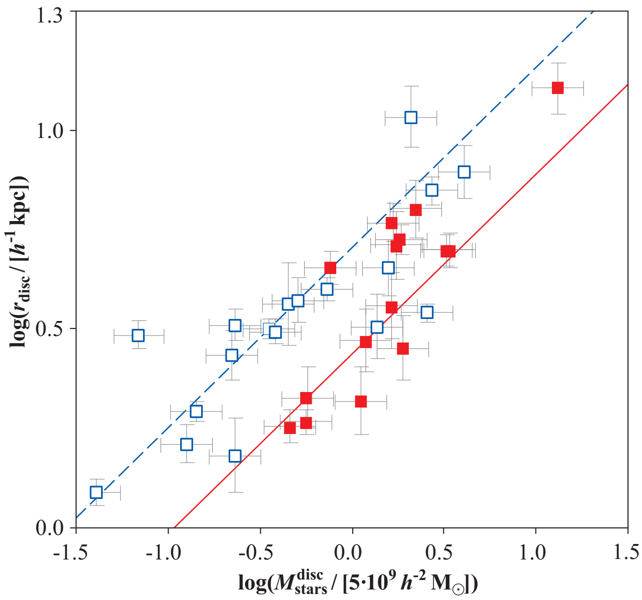

Figure B2. Relation between disc scale radius $r_{\rm disc}$ and disc stellar mass $M_{\rm stars}^{\rm disc}$. Squares represent 34 nearby spiral galaxies observed by Kregel et al. (2002). Filled squares correspond to Sa/Sb-type galaxies, and empty squares represent Sc/Sd-type galaxies. These data were used to fit the free parameters of the model in Eq. (24). The solid line shows this model for $T = 2$ (i.e. Sab-type galaxies), while the dashed line shows the model for $T = 6$ (i.e. Scd-type galaxies). The slope of these power-laws is $\alpha_2 = 0.45 \pm 0.05$, consistent with a Freeman law (McGaugh et al. 1995).

APPENDIX B: DIVERSE PHENOMENOLOGICAL RELATIONS

This section summarizes the two phenomenological relations given in Eqs. (23) and (24).

B1 Stellar mass versus gas mass

From the galaxy sample presented in Appendix A, we extracted all 25 Scd/Sd-type galaxies ($6 \leqslant T \leqslant 9$), that is all objects approximating pure discs. For these objects the total gas masses $M_{\rm gas}$ were calculated via $M_{\rm gas} = (M_{\rm HI} + M_{\rm H_2})/\beta$. Additionally, we estimated the stellar mass $M_{\rm stars}$ of each galaxy from the I-band magnitude $M_{\rm I}$ via (Mo et al. 1998),

$$\log(M_{\rm stars}/{\rm M}_\odot) = 1.66 + \log(\Upsilon_{\rm I}) - M_{\rm I}/2.5, \qquad \text{(B1)}$$

where the mass/light-ratio $\log(\Upsilon_{\rm I}) = 1.2$ has been adopted from McGaugh & de Blok (1997).

The resulting data points displayed in Fig. B1 reveal an approximate power-law relation between $M_{\rm gas}$ and $M_{\rm stars}$. We have fitted the corresponding free parameters α_1, γ_1 to the data points by minimizing the x-y-weighted rms-deviations. 1-σ errors for these parameters were obtained via a bootstrapping method that uses 10^3 random half-sized subsamples of the 25 galaxies and determines the power-law parameters for every one of them. The standard deviations of the distributions for α_1 and γ_1 are then divided by $\sqrt{2}$ to estimate 1-σ confidence intervals for the full data set. The best power-law fit and its 1-σ confidence interval are displayed in Fig. B2, while explicit numerical values are given in Section 5.3.

To first order, one would expect that $M_{\rm stars}$ depends linearly on $M_{\rm gas}$, if both masses scale linearly with the mass of the parent haloe. The over-proportional growth of $M_{\rm stars}$ ($\alpha_1 = 1.46 \pm 0.10 > 1$) could be explained by the fact that more massive galaxies are generally older and therefore could convert a larger fraction of hydrogen gas into stars.

B2 Scale radius versus stellar mass

Kregel et al. (2002) investigated a sample of 34 nearby edge-on spiral galaxies, drawn from the ESO-LV catalog (Lauberts & Valentijn 1989) using four selection criteria: (i) inclination $i \geqslant 87$ deg, (ii) blue diameter $D_{25}^{\rm B} > 2.2$ arcmin, (iii) Hubble type from S0–Sd, (iv) only regular field galaxies, i.e. no interacting systems, no warped or lopsided systems. This sample is complete in terms of sample selection (see Kregel et al. 2002; Davies 1990), but the sample volume is too small to contain rare objects. For each galaxy in the sample Kregel et al. (2002) determined the scale radius $r_{\rm disc}$ of the stellar disc from the I-band luminosity profiles. They also obtained the morphological Hubble type T for each source from the Lyon/Meudon Extragalactic Database (LEDA). Additionally, we estimated the disc stellar masses $M_{\rm stars}^{\rm disc}$ from the I-band magnitudes of the disc components according to Eq. (B1).

Using these data, we investigated the relations between $M_{\rm stars}^{\rm disc}$, $r_{\rm disc}$ and T. The data points shown in Fig. B2 suggest the approximate power-law with Hubble type correction of Eq. (24). The best fitting parameters α_2, γ_2, and δ were obtained as in Section B1 and explicit numerical values with errors are given in Section 5.3.

To first order, the $r_{\rm disc}$–$M_{\rm stars}^{\rm disc}$ relation can be understood in terms of a dark matter halo with an isothermal, singular, and spherical structure (e.g. Mo et al. 1998). This model predicts that the virial radius $r_{\rm vir}$ is proportional to the cubic root of the dark matter mass $M_{\rm DM}$ at any fixed cosmic time. If $r_{\rm disc}$ were proportional to $r_{\rm vir}$ and $M_{\rm stars}^{\rm disc}$ were proportional to $M_{\rm DM}$, one would expect $r_{\rm disc}$ to scale as $(M_{\rm stars}^{\rm disc})^{1/3}$. Our empirical result, $\alpha_2 = 0.45 \pm 0.05$, shows a slightly stronger scaling, consistent with the empirical Freeman law ($\alpha_2 = 0.5$), according to which disk galaxies have approximately constant surface brightness (McGaugh et al. 1995).

The secondary dependence of the $r_{\rm disc}$–$M_{\rm stars}^{\rm disc}$ relation on the

Hubble type T probably has multiple reasons: (i) early-type galaxies have more massive stellar bulges, which present an additional central potential that contracts the disc; (ii) bulges often form from disc instabilities, occurring preferably in systems with relatively low angular momentum, and hence early-type galaxies are biased towards smaller angular momenta and smaller scale radii; (iii) larger bulges, such as the ones of lenticular and elliptical galaxies, often arise from galaxy mergers, which tend to reduce the specific angular momenta and scale radii (see also Obreschkow et al. 2009).